AD-A039 545 STANFORD RESEARCH INST MENLO PARK CALIF F/G 15/3
PREDICTION OF CONGREGATE-CARE SPACE IN NON-METROPOLITAN COUNTIE--ETC(U)
FEB 77 W E STROPE, B J NEITZEL DCPA01-76-C-0298
UNCLASSIFIED SRI-CRESS-22 NL

1 OF 1
AD A039545

END

DATE FILMED
6-77

12

Final Report
CRESS–22

February 1977

PREDICTION OF CONGREGATE-CARE SPACE IN NON-METROPOLITAN COUNTIES

CONTRACT DCPA01–76–C–0298
DCPA Work Unit 2312G

STANFORD RESEARCH INSTITUTE
Menlo Park, California 94025 · U.S.A.

STANFORD RESEARCH INSTITUTE
SRI-Washington D.C. Office · U.S.A.

Final Report
CRESS–22

February 1977

PREDICTION OF CONGREGATE-CARE SPACE IN NON-METROPOLITAN COUNTIES

By: WALMER E. STROPE
BETTY J. NEITZEL

Prepared for:

DEFENSE CIVIL PREPAREDNESS AGENCY
WASHINGTON, D.C. 20301

CONTRACT DCPA01–76–C–0298
DCPA Work Unit 2312G

DCPA REVIEW NOTICE

This report has been reviewed in the Defense Civil Preparedness Agency and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the Defense Civil Preparedness Agency.

1611 NORTH KENT STREET, ROSSLYN PLAZA, ARLINGTON, VIRGINIA 22209 • TELEPHONE (703) 524-2053
Main Office and Laboratories: Menlo Park, California 94025, U.S.A.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
(6) PREDICTION OF CONGREGATE-CARE SPACE IN NON-METROPOLITAN COUNTIES.

5. TYPE OF REPORT & PERIOD COVERED
(9) FINAL rept.,

(14) SRI

6. PERFORMING ORG. REPORT NUMBER
CRESS-22

7. AUTHOR(s)
(10) Walmer E. Strope
Betty J. Neitzel

8. CONTRACT OR GRANT NUMBER(s)
(15) DCPA01-76-C-0298 new

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Stanford Research Institute
333 Ravenswood Avenue
Menlo Park, California 94025

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
DCPA Work Unit 2312G

11. CONTROLLING OFFICE NAME AND ADDRESS
Defense Civil Preparedness Agency
Washington, D.C. 20301

12. REPORT DATE
(11) Feb[illegible]77

13. NO. OF PAGES
94

15. SECURITY CLASS. (of this report)
Unclassified

(12) 83p.

14. MONITORING AGENCY NAME & ADDRESS (if diff. from Controlling Office)

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
N/A

16. DISTRIBUTION STATEMENT (of this report)
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
CIVIL DEFENSE
HOST AREA SURVEY
PREDICTION
FACILITY

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
A procedure is presented for predicting the amount of congregate-care space existing in a non-metropolitan county prior to a facility survey. Estimates of accuracy and reliability are given, based on a 60-county sample from the 1975 DCPA host area survey. Suggestions are made for improvement. The procedure is not suitable for predicting the outcome of a partial survey.

332 500

DD FORM 1 JAN 73 1473
EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19. KEY WORDS (Continued)

20 ABSTRACT (Continued)

CRESS-22
Project EGU-5583

PREDICTION OF CONGREGATE-CARE SPACE IN NON-METROPOLITAN COUNTIES

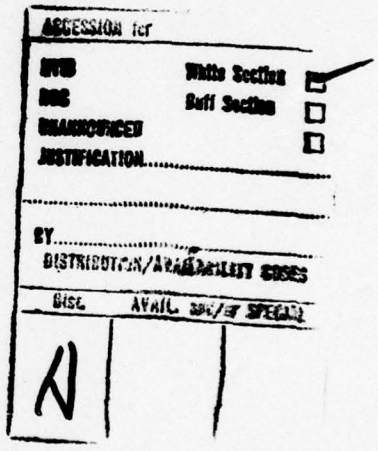
Final Report
February 1977

by

Walmer E. Strope

and

Betty J. Neitzel



for

Defense Civil Preparedness Agency
Washington, D.C. 20301

Contract DCPA01-76-C-0298
DCPA Work Unit 2312G

Approved for Public Release;
Distribution Unlimited

DCPA Review Notice

This report has been reviewed in the Defense Civil Preparedness Agency and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the Defense Civil Preparedness Agency.

STANFORD RESEARCH INSTITUTE
1611 North Kent Street
Arlington, Virginia 22209

PREDICTION OF CONGREGATE-CARE SPACE IN NON-METROPOLITAN COUNTIES

SUMMARY

A procedure has been developed for predicting, in advance of a facility survey, the amount of congregate-care space to be found in whole non-metropolitan counties in the United States. Based on a 60-county sample from the 1975 host area survey, the prediction technique provides an unbiased estimate of per capita congregate-care space with a standard deviation of about 18 percent.

The procedure consists of assigning the county of interest an initial estimate of 3.10 spaces per capita, which is then adjusted upward or downward on the basis of a comparison of certain census data for the county with the national average for non-metropolitan counties. In addition, separate estimates are made for unique congregate-care resources for which no adequate census indicator has been found. These include special facilities, such as mines, caves, and tunnels, unusually large industrial facilities, private colleges, and seasonal tourist facilities. Information for these estimates must be obtained from one familiar with the county.

The estimate using only census indicators can be computerized. It tends to underpredict resource-rich counties but can be adjusted to give an unbiased estimate with a standard deviation of about 25 percent.

The principal source of prediction error is believed to be the uncertainties in the interpretation of the 1975 survey results. Suggestions are made for reducing these uncertainties.

The prediction technique is not suitable for predicting the outcome of a survey of a part of a county, including the non-risk part of a metropolitan county. In general, the technique overpredicts the amount of congregate-care space to be found in the non-urbanized parts of metropolitan counties.

CONTENTS

ILLUSTRATIONS

TABLES

ABSTRACT

A procedure is presented for predicting the amount of congregate-care space existing in a non-metropolitan county prior to a facility survey. Estimates of accuracy and reliability are given, based on a 60-county sample from the 1975 DCPA host area survey. Suggestions are made for improvement. The procedure is not suitable for predicting the outcome of a partial survey.

I INTRODUCTION

Background

Government policy in Crisis Relocation Planning (CRP) is that people relocated from areas presumed at risk will be housed in non-residential, non-farm buildings and other suitable facilities, called congregate-care space (CCS) in host counties. The Defense Civil Preparedness Agency (DCPA) has accomplished surveys of congregate-care resources in selected host counties during the summers of 1974, 1975, and 1976 and presumably will continue to devote resources to this survey effort. The level of effort available, however, dictates that the approximately 3000 counties in the country will be surveyed over a considerable number of years.

In SRI's investigation of the feasibility of relocating the urban population of the Northeast Corridor during a crisis,[1] it was found that use of an average estimate of hosting capacity in planning the allocation of relocatees would in many cases result in assignments that would need wholesale revision once the hosting resources in the various host counties had been identified through a survey. The difficulty is that actual congregate-care space, which appears to be the primary measure of hosting capacity, is known to vary widely from the average. The problem is compounded in the Northeast because the numbers of people to be hosted are so large that a space allocation of about 20 square feet (1.86 square meters) per person will be necessary in lieu of the peacetime emergency housing standard of 40 square feet (3.72 square meters). This reduces the amount of error that can be tolerated in the knowledge of housing resources. The situation is even more constrained in California where an allocation of 10 square feet (0.93 meters) may be necessary. This allocation is the same as the space allocation in fallout shelter.

SRI analyzed the 1974 host area survey results as part of Contract No. DCPA01-74-C-0293. Data for 28 non-metropolitan counties were found suitable for this analysis. The number of 40-square-foot housing spaces per host resident (per capita CCS) ranged from a high of eight to a low of just under two; the average value was 3.7. Linear regression analysis showed some promise of providing a satisfactory prediction method based on resident population and some economic indicators, such as retail sales. The analysis also showed some promise that methods could be developed to predict the capacity of certain important building types, such as schools and commercial buildings, from readily available census data. Thus, it seemed possible that more analysis and better survey data could produce a prediction method that,

if used, could allow State and regional planning to proceed in advance of the actual survey of housing resources.

Further analysis of the 1974 survey data was accomplished as part of the Northeast Corridor feasibility study. The survey data were divided into four categories and compared with information published by the Bureau of the Census. Space in buildings housing population-oriented activities, such as education, religion, government, public services, and amusement, were assumed to be related to county population. Commercial space was compared to retail sales. Space in industrial buildings was compared to manufacturing employment. Space in facilities serving a specialized segment of the population, such as barracks, dormitories, correctional institutions, and the like, were compared to census data on the proportion of the population living in group quarters. The results of this "Four-Element Method" were compared to the prior estimates based on applying an average per capita CCS to the county population. Comparison for 28 counties showed that estimates for 18 were improved and for 10 were degraded. There was, however, a decrease in very large errors (over 50 percent).

During the summer of 1975, parts or all of some 200 counties were surveyed for congregate-care capacity. The work reported here concerned use of these data in an attempt to develop a more satisfactory method for estimating the potential congregate-care space in counties where a survey has not yet been accomplished.

Objectives and Scope

The objectives of this work, as specified in Contract No. DCPA-01-76-C-0298, are:

> The Contractor, in cooperation and consultation with the Government, shall furnish the necessary personnel, facilities, materials, and such other services as may be required to continue to develop procedures for estimating shelter (congregate care) capacity in host areas.
>
> The Contractor shall perform specific work and services as follows:
>
> (1) Analyze 1974 and 1975 host area survey data and compare the developed technique with this data for an estimate of accuracy and reliability of the technique.
>
> (2) Identify areas of inconsistency between data and the predictive technique and make recommendations for the resolution of the inconsistency.

Purpose

The purpose of this report is to present a predictive technique for estimating congregate-care capacity that is significantly more accurate and reliable than that previously available, to report the analysis that led to the development of the method, and to identify areas of inconsistency between the data and the predictive technique that require resolution.

Organization of the Report

There are six sections to this report, including this introduction. Section II presents the predictive technique in a form suitable for manual assessment of the per capita CCS likely to be found in a particular county. The next section presents the estimates made by this technique for surveyed counties and compares them with the survey results. Section IV describes the analysis that led to the technique, including a discussion of procedures that were discarded. Section V discusses the prediction data base; that is, the 1975 survey results and adjustments made for purposes of analysis. The final section presents our conclusions and recommendations.

II THE PREDICTIVE TECHNIQUE

A recommended procedure for predicting the amount of congregate-care space to be found in non-metropolitan counties in the United States is presented in the following pages. The chief characteristics of the method are:

(1) The calculation is done in terms of per capita spaces; that is, the number of 40-square feet housing spaces per host county resident.

(2) Each county is assigned an initial quota of 3.1 spaces per capita. This initial level is then adjusted either higher or lower according to simple rules.

(3) The primary data for adjustment of the initial level comes from the Bureau of the Census publication, County and City Data Book, 1972.[2] This version of census data must be used, as the multipliers and national averages used in the calculation are taken from this source.

(4) Four items of data from the above source are needed. These items are used in pairs in the calculation. The first two are economic in character: per capita income and retail sales. The second two are activity-oriented: percent employed in government and percent employed in service industries. These indices for a particular county are compared with the national average for non-metropolitan counties.

(5) Variations from the national average are multiplied by a weighting or conversion factor into units of per capita CCS. The conversion factors are twice as large for below-average counties as they are for above-average counties. Simple rules are provided to choose which results, if any, are to be added to or subtracted from the initial level of per capita CCS. These rules provide for counties with mixed economic indicators or with work force characteristics that vary widely from the average.

(6) The application of the census data results in a modified per capita CCS that was found to be within plus or minus 25 percent of the survey result in 77 percent of a sample of 60 non-metropolitan counties surveyed in 1975. A final portion of the calculation, which is based on information acquired in the county itself, allows inclusion of per capita CCS in facilities not covered by the census indicators. These include mines and caves, unusually large industrial plants, private universities, and summer resort facilities. This portion corrects underestimates in housing-rich counties.

A form for hand calculation of a predicted per capita CCS for a non-metropolitan county, together with needed instructions, is presented in the following pages. The calculation is, of course, readily performed on data-processing equipment and the 1972 County and City Data Book (CCDB) is available in machine-readable form. Hence, estimates based on the census data could be done by computer, leaving only the addition of other resources as a task in the field.

The hand-calculation form is similar to an income-tax form. The adjustments to the initial level are done in an Estimate Summary. The calculation of the adjustments is performed in three "schedules", one for the economic adjustment, one for the activity adjustment, and one for the additional resources that may exist in the county. Thus, the schedules are completed first and the results brought forward to the Estimate Summary to produce the final estimate or prediction.

CONGREGATE-CARE SPACE (CCS) ESTIMATING FORM FOR NON-METROPOLITAN COUNTIES

County Name_____ RSAC No._____ State_____

ESTIMATE SUMMARY	Per Capita
Line 1: Initial Estimate	+ 3.10
Line 2: Economic Adjustment (from Schedule A)	
Line 3: Activity Adjustment (from Schedule B)	
Line 4: Additional Resources (from Schedule C)	+____
Line 5: Final Estimate of CCS (See Instruction 1)	+

SCHEDULE A: ECONOMIC ADJUSTMENT

Line 1: Per Capita Money Income (CCDB, Table 2, Col. 67) .	$
Line 2: Average Per Capita Money Income	$2480
Line 3: Excess (+) or Deficiency (-) (Line 1 less Line 2)	$

If Line 3 is +, multiply by 0.001 and enter on Line 4 as increase (+).
If Line 3 is -, multiply by 0.002 and enter on Line 4 as decrease (-).

Line 4: Potential Money Income Adjustment	
Line 5: Retail Sales (CCDB, Table 2, Col. 135) . . . . . (See Instruction 2)	$
Line 6: Book Population (CCDB, Table 2, Col. 3)	
Line 7: Per Capita Retail Sales (Line 5 ÷ Line 6)	$
Line 8: Average Per Capita Retail Sales	$1350
Line 9: Excess (+) or Deficiency (-) (Line 7 less Line 8).	$

If Line 9 is +, multiply by 0.001 and enter on Line 10 as increase (+).
If Line 9 is -, multiply by 0.002 and enter on Line 10 as decrease (-).

Line 10: Potential Retail Sales Adjustment . . . . . .

If Line 4 and Line 10 are both increases (+), enter the largest increase on Line 11 and on Line 2 of the Estimate Summary.

If Line 4 and Line 10 are both decreases (-), enter the largest decrease on Line 11 and on Line 2 of the Estimate Summary.

If Line 4 and Line 10 are <u>not</u> both increases or both decreases, enter zero on Line 11 and on Line 2 of the Estimate Summary.

Line 11: Economic Adjustment . . . . . (See Instruction 3)

SCHEDULE B: ACTIVITY ADJUSTMENT

Line 1: Government Employment (CCDB, Table 2, Col. 44) . . . %
Line 2: Average Government Employment 16.3%
Line 3: Excess (+) or Deficiency (-) (Line 1 less Line 2) %

If Line 3 is +, multiply by 0.05 and enter on Line 4 as increase (+).
If Line 3 is -, multiply by 0.10 and enter on Line 4 as decrease (-).

Line 4: Government Activity Adjustment . . . . . .

Line 5: Employment in Services (CCDB, Table 2, Col. 41) %
Line 6: Average Employment in Service Industries 7.0%
Line 7: Excess (+) or Deficiency (-) (Line 5 less Line 6) %

If Line 7 is +, multiply by 0.10 and enter on Line 8 as increase (+).
If Line 7 is -, multiply by 0.20 and enter on Line 8 as decrease (-).

Line 8: Service Activity Adjustment . . . . . .

Line 9: Gross Activity Adjustment (Line 4 plus Line 8) .

Line 10: Percent Work Outside County (CCDB, Table 2, Col. 49) %

If Line 9 is +, and Line 10 is less than 24%, enter Line 9 increase on Line 11 and on Line 3 of the Estimate Summary.

If Line 9 is +, and Line 10 is 24% or more, enter zero on Line 11 and on Line 3 of the Estimate Summary.

If Line 9 is -, and Line 10 is 8% or more, enter Line 9 decrease on Line 11 and on Line 3 of the Estimate Summary.

If Line 9 is -, and Line 10 is less than 8%, enter 50% of Line 9 decrease on Line 11 and on Line 3 of the Estimate Summary.

Line 11: Net Activity Adjustment . . . . . .

SCHEDULE C: ADDITIONAL RESOURCES

Line 1: Book Population (from Schedule A, Line 6) . . .
Line 2: Multiply Line 1 by 0.10 . . . . . .

Line 3: Does county contain special facilities (See Instruction 4) with probable space in excess of Line 2? YES____ NO____

Line 4: If Line 3 is yes, estimate of total floor space sq. ft.
Line 5: Divide Line 4 by 40 if not zero . . . . . spaces
Line 6: Divide Line 5 by Line 1 . . . . . per capita spaces
Line 7: Does county contain major industrial plants (see Instruction 5) with probable spaces in excess of Line 2? YES____ NO____

SCHEDULE C: ADDITIONAL RESOURCES (Cont'd.)

Line 8:	If Line 7 is yes, estimate of large facility floor area	sq. ft.
Line 9:	Divide Line 8 by 75 if not zero	spaces
Line 10:	Divide Line 9 by Line 1	per capita spaces

Line 11: Does county contain one or more <u>private</u> colleges or universities (See Instruction 6) with probable spaces in excess of Line 2? YES____ NO____

Line 12:	If Line 11 is yes, estimate of total floor space	sq. ft.
Line 13:	Divide Line 12 by 50 if not zero	spaces
Line 14:	Divide Line 13 by Line 1	per capita spaces

Line 15: Does county have significant <u>seasonal</u> resort facilities available to the public (See Instruction 7) with probable spaces in excess of Line 2? YES____ NO____

Line 16:	If Line 15 is yes, estimate of additional floor space	sq. ft.
Line 17:	Divide Line 16 by 50 if not zero	spaces
Line 18:	Divide Line 17 by Line 1	per capita spaces

If Line 8, Schedule B, is negative, enter Line 18 total on Line 19.

If Line 8, Schedule B, is positive, add it to 0.7, subtract from Line 18 and if difference is positive, enter on Line 19. Otherwise, enter zero on Line 19.

Line 19:	Seasonal resort facilities	per capita spaces
Line 20:	Additional Resources (Add Lines 6, 10, 14 and 19 and enter here and on Line 4 of the Estimate Summary . .	per capita spaces

INSTRUCTIONS

<u>Instruction 1: Estimate of Per Capita CCS</u>. The estimate of per capita congregate-care spaces available in the county may be multiplied by the population of the county to obtain an estimate of the gross number of 40-square feet spaces that might be expected in an actual survey of nonresidential, non-farm facilities. Since a portion of this space

will be in facilities that may prove unsuitable for housing people or that may be needed for essential activities, use two-thirds of the gross number as the net spaces available. If a reduced space allocation must be used in the planning region to accommodate the risk population within reasonable travel distances, multiply the resulting net figure by the ratio of the standard 40 square feet to the reduced allocation.

Note that the Final Estimate is based on adjustments made to an initial assignment of 3.1 CCS per host-county resident. This figure is about 10 percent less than the average for non-metropolitan counties. In past surveys, about half of surveyed counties were found to contain facilities with gross CCS within plus or minus 25 percent of the average. However, the full range of variation is from about 3 times the average to only 1/3 the average.

The adjustments summarized in Lines 2 and 3 of the Estimate Summary are based on census data in the 1972 County and City Data Book issued by the Bureau of the Census. This issue must be used if a valid estimate is to be made. Other than this restriction, the economic and activity adjustments of Schedules A and B can be made with no personal knowledge of the county. These adjustments can be positive or negative; that is, increases to or deductions from the initial estimate of 3.1. It is very important to keep track of these increases and decreases by using the proper sign (+ or -) and to indicate on Lines 2 and 3 of the Estimate Summary by the proper sign whether the adjustment is an increase or a decrease in the per capita CCS.

If only the adjustments that can be made from use of the 1972 County and City Data Book are made (Lines 2 and 3 but not Line 4) the likelihood that the survey result will be within plus or minus 25 percent of the "desk-top" estimate is increased to about 75 percent. In particular, failure to execute Schedule C will underestimate the per capita CCS in counties rich in resources not reflected adequately in the census indicators. Line 4 of the Estimate Summary is always an increase in the per capita CCS when it is not zero. To execute Schedule C, the planner must have personal knowledge of additional resources in the county or must obtain the required information from county officials and State agencies as described in subsequent instructions. If all elements of the Estimate Summary are completed, the likelihood that the survey result will be within plus or minus 25 percent of the Final Estimate is increased to about 85 percent and the likelihood that the error is greater than about 35 percent is quite small.

Instruction 2: Retail Sales. The retail sales figure in Column 135 of the county table (Table 2) of the 1972 County and City Data Book is in thousands of dollars, as indicated at the head of the column. Therefore, the planner must add three more zeros to the number given to obtain the appropriate value for entry in Line 5. Otherwise, when

divided by the "book population" on Line 6, the per capita retail sales will be a thousand times too small. As a check, note that the average per capita retail sales in non-metropolitan counties is $1350 (Line 8). Only rarely will the per capita retail sales for a particular county fall below $1000 or over $3500. Note also that it is important to use the book population on Line 6. Do not use an updated or corrected population figure, as the conversion factors used to fill in Line 10 are keyed to the population listed in Column 3 of Table 2.

Instruction 3: Economic Adjustment. The economic adjustment is based on comparison of two factors, Money Income and Retail Sales, with the national averages for non-metropolitan counties. The weighting or conversion factors that determine the imputed effect on facility space are twice as large for deficiencies (below-average counties) as they are for counties that are above average. Neither measure by itself is an adequate indicator of the facility space generated by economic activities. If both factors are above average, a strong resource is predicted and the larger of Lines 4 and 10 should be entered here and on Line 2 of the Estimate Summary. Make sure the entry is labeled + as an additive adjustment. Similarly, if both factors are below average, a weak resource is predicted and the most negative (larger of the minus values) should be used. In many counties, one factor may be above average while the other is below average. For example, counties containing a large college or university often show a below-average money income (because of the students) and an above-average per capita retail sales. Counties having a larger commercial center in a neighboring county may have above-average money income and below-average retail sales. In these cases, the data indicate that it is best to regard the county as average economically and to enter no economic adjustment. If an economic adjustment is indicated according to the above rules, make sure that the positive or negative sign is used to indicate whether it should be added to or subtracted from the initial estimate.

Instruction 4: Special Facilities. One kind of housing resource that is not accounted for by the census indicators in Schedules A and B is the space that may be available in what are called "special facilities." Special facilities are defined by DCPA as the following: (1) Mines, (2) Caverns or caves, (3) Tunnels, (4) Subways, (5) Underpasses, (6) Underground storage facilities, (7) Inactive military works, and (8) Other special facilities. This set of designations was intended to be applied to shelter from fallout but many may be suitable for temporary housing as well.

If the county is known to contain a number of mines or caves, it must be determined whether parts of them are suitable for temporary habitation. That is, would they be surveyed for this purpose? Large tunnels may also be considered. Subways are not found in non-metropolitan counties. Underground storage facilities might exist for potatoes or other crops. Inactive military works may be an important

resource in some counties. The definition should be broadened to include any inactive military or government installation that would not be reflected in the measure of government employment in Schedule B. Among "other" facilities that have been considered for survey are highway culverts.

If the county may contain any special facilities, a knowledgeable local official should be asked to judge whether any are usable and whether they are likely to hold more people at 40 square feet per person than the number on Line 2. If not, their contribution would be too small to encourage further consideration. Thus, a single facility in a county of modest population may be worth pursuing, whereas many large facilities would be needed to make a significant per capita contribution in a county with a large population. When the contribution is likely to be significant, arrangements should be made to get a reasonable estimate of total usable floor space, short of an actual survey. In addition to local sources of information, State agencies concerned with mining, geology, transportation, agriculture, and military affairs may be of assistance. Once an approximation of the total floor area available is entered into Line 4, it is divided by 40 to obtain congregate-care spaces and then by the book population to obtain the per capita spaces predicted prior to survey.

Instruction 5: Industrial Facilities. The economic indicators employed in Schedule A provide a measure of industrial as well as commercial and tax-supported facilities that might be in the county. In the average non-metropolitan county, about 0.25 congregate-care spaces are found in industrial facilities and this resource, which is usually composed of a number of locations, is reflected in the initial estimate in the Estimate Summary. However, if the county has one or more unusually large industrial plants, this resource will be undercounted in the average figure. How large a plant must be to be considered an additional resource depends upon the county population. In a county of only a thousand or so persons, a single cotton gin or processing plant may contain 50,000 square feet of usable floor area and, hence, more than one space for every resident. In more populous counties, a major industrial park or fabricating plant may qualify. Comparison should be made with similar counties known to the planner in determining whether any industrial facilities should be counted as an additional resource. Since a survey of many or all industrial facilities is not intended, the names and locations of major facilities should be readily obtained from a knowledgeable local official. As discussed in Instruction 4, a preliminary estimate of the probable number of spaces available in a specific plant site should be obtained before going further. This information should be compared with the number on Line 2. If a single plant site is unlikely to provide at least one-tenth space per capita, it should not be considered an additional resource unless there are several such sites. If the answer to Line 7 is yes, then the total floor space available should be obtained from the facility management and entered on Line 8. Since

industrial facilities are usually occupied in considerable part by nonmovable machinery and equipment, the estimate of floor area should be divided by 75 on Line 9 to obtain a prediction of housing spaces. Line 9 is then divided by the population of the county to obtain the prediction of per capita spaces.

Instruction 6: Private Colleges. Most institutions of higher learning in non-metropolitan counties are supported and operated by some level of government. The amount of government employment in the county considered in Schedule B will be a sufficient measure of the space in such institutions. Large private colleges and universities, such as Dartmouth in New Hampshire or St. Leo in Florida will not be counted by this means. Therefore, the planner should establish whether one or more private residence institutions exist in the county with substantial potential capacity. Where these are found, an estimate of floor space should be obtained from the institution administration. The calculations to obtain predicted per capita spaces are similar to those for special and industrial facilities. There also may be parochial or private schools below the college level which have more than the normal number of school buildings on their property. For example, preparatory schools have residence buildings and these should also be included in the estimate. In many areas, high schools, both public and private, may have separate buildings for gymnasiums. This space is already accounted for in the initial estimate and so these schools should not be considered to be additional resources in this section.

Instruction 7: Resort Facilities. The amount of service employment in the county considered in Schedule B is intended to measure congregate-care space in hotels, motels, camps and allied supporting services for non-residents of the county. A weakness of this measure is that the census information is obtained during early April. This time of year is generally the off-season tourist period. Therefore, it will seriously undercount summer resort areas, such as Mackinac Island, Michigan, where employment is seasonal and often transient. It is also possible that winter resort areas will be undercounted as in some locations the peak seasonal activity may be over by mid-March. If the county has extensive resort facilities (not merely private vacation homes or cottages), they may be an additional resource above and beyond the space accounted for in the initial estimate. In the average county, hotel and motel spaces account for about 0.4 space per capita and other supporting services about 0.3 spaces. Hence resort facilities would need to contribute at least one space per capita to be considered excessive and the contribution of the Service Activity Adjustment (Schedule B, Line 8) must be considered as well. Nonetheless, there are a substantial number of counties that will qualify, including a low-population county in Nevada having a single hotel-casino with space for twice the county population! On Line 16, make sure to estimate the additional floor space provided by resort facilities. If Line 8 of Schedule B is negative, essentially all motel, hotel, and camp space available to the

public can be included. If Line 8 is positive, the amount should be added to 0.7 spaces. Only spaces on Line 18 in excess of this number should be considered. The local Chamber of Commerce or motel-owners associations are good sources of information.

III COMPARISON WITH SURVEY DATA

In this section, the results obtained with the predictive procedures of Section II will be compared with available survey results. These comparisons form the basis for estimating the accuracy and reliability of the predictive technique. Emphasis will be placed on the results of the 1975 summer survey for several reasons. First, certain of these results were used in the development of the technique. Second, the data available from the 1974 survey are insufficient to permit execution of Schedule C of the procedure. Finally, it is believed that the survey procedures used in the summer of 1975 were considerably improved over those used in 1974. (The survey results for the summer of 1976 are not yet available.)

1975 Complete Counties

The predictive technique of Section II attempts to estimate the per capita amount of congregate-care space that would be found in a non-metropolitan county if it were surveyed in a manner similar to the 1975 survey. Most of the nearly 200 counties surveyed during the summer of 1975 were surveyed only in part. Since we are interested in the relationship of the survey results to the resident population (per capita CCS), the determination of the "survey *population*" *in partially-surveyed* counties becomes an additional source of error. Therefore, the survey was searched for those counties that appeared to have been completely surveyed. The methods used to evaluate whether a county had been completely surveyed are described in Section V.

A total of 60 counties were identified as probably completely surveyed. The performance of the technique of Section II against the survey results for this group is shown in Table 1. The counties are ordered by survey result from highest per capita CCS to lowest. Esmeralda County, Nevada, has the highest per capita CCS (9.10) and Bienville Parish, Louisiana, has the lowest (1.28). The highest 30 counties are on the first page of Table 1; the lowest 30 counties are on the second page.

The headings of the columns in Table 1 refer to the format given in Section II. "S1" refers to Line 1 of the Estimate Summary, the Initial Estimate, and is the same for all counties. "A4" and "A10" refer to Lines 4 and 10 of Schedule A, which show the potential money income and retail sales adjustments. "S2" is the economic adjustment brought forward to Line 2 of the Summary according to the rules of Schedule A. "B4" and "B8" refer to the government and service activity adjustments on Lines 4 and 8 of Schedule B. "S3" indicates the net activity adjustment brought forward to Line 3 of the Summary. The symbol # indicates cases where an unusual deviation in the percent working outside the county caused a modification of the gross activity adjustment according to the rules of Schedule B.

Table 1

PREDICTIONS FOR 60 NON-METROPOLITAN COUNTIES

Line---→ / County, State	S1	A4	A10	S2	B4	B8	S3	C6	C10	C14	C18	S5	Survey Result	Error
Esmeralda, NV	3.1	+0.97	-2.01	--	+0.41	+0.54	+0.95	+3.00	--	--	+2.00	9.05	9.10	+1%
Nye, NV	3.1	+1.36	+0.24	+1.36	+0.44	+1.24	+1.68	+1.25	--	--	--	7.39	8.80	+19%
Collier, FL	3.1	+1.68	+0.06	+1.68	-0.51	+0.66	+0.15	--	--	--	--	4.93	5.78	+17%
Albany, WY	3.1	+0.26	-0.12	--	+1.29	+0.10	+1.39	--	--	--	--	4.49	5.57	+24%
Lincoln, NV	3.1	+0.08	-0.40	--	+1.18	-0.14	+1.04	+0.85	--	--	--	4.99	5.03	+1%
Whitman, WA	3.1	+0.30	-0.15	--	+1.33	-0.22	+1.11	--	--	--	--	4.21	4.99	+19%
Luce, MI	3.1	-0.34	-0.47	-0.47	+1.51	-0.16	+1.35	--	--	--	+0.60	4.58	4.78	+4%
San Luis Obispo, CA	3.1	+0.38	+0.03	+0.38	+0.64	+0.28	+0.92	--	--	--	--	4.40	4.74	+8%
Eastland, TX	3.1	-0.37	+0.16	--	-0.31	+0.14	-0.09#	--	+0.50	--	+0.20	3.71	4.71	+27%
Box Elder, UT	3.1	-0.30	-0.05	-0.30	+0.62	-0.26	+0.36	--	+1.00	--	--	4.16	4.38	+5%
Baylor, TX	3.1	+0.04	+0.83	+0.83	-0.32	+0.54	+0.22	--	--	--	--	4.15	4.27	+3%
Manatee, FL	3.1	+0.37	+0.03	+0.37	-0.28	+0.41	+0.13	--	+0.30	--	--	3.90	4.21	+8%
Latah, ID	3.1	+0.17	+0.27	+0.27	+1.09	-0.06	+1.03	+0.10	--	--	--	4.50	4.14	-8%
Mason, TX	3.1	-0.88	+0.60	--	-0.21	+0.08	-0.07#	--	+0.25	--	--	3.29	3.91	+19%
Knox, TX	3.1	-1.08	+0.07	--	+0.04	+0.08	+0.12	+0.10	--	--	--	3.32	3.86	+16%
Polk, FL	3.1	+0.09	+0.12	+0.12	-0.42	+0.16	-0.13#	--	+0.30	+0.20	--	3.59	3.82	+6%
Yuba, CA	3.1	-0.22	+0.07	--	+0.39	+0.16	+0.55	--	--	--	--	3.65	3.58	-2%
Dickens, TX	3.1	-0.58	+0.05	--	+0.10	+0.04	+0.14	--	+0.10	--	--	3.34	3.55	+6%
King, TX	3.1	+0.41	-1.08	--	+0.01	-0.84	-0.42#	--	--	--	??	2.68	3.55	+32%
Lincoln, WA	3.1	+0.73	+0.45	+0.73	+0.15	+0.02	+0.17	--	--	--	--	4.00	3.51	-12%
Hood, TX	3.1	+0.10	-0.02	--	-0.10	+0.04	-0.06	--	--	--	--	3.04	3.47	+14%
Haskell, TX	3.1	-0.58	-0.13	-0.58	-0.50	+0.14	-0.18#	--	+0.10	+0.10	--	2.54	3.43	+35%
Hardeman, TX	3.1	+0.01	-0.03	--	-0.07	+0.24	+0.17	--	+0.45	--	--	3.72	3.42	-8%
Ogemaw, MI	3.1	-0.46	+0.10	--	+0.02	-0.40	-0.38	--	+0.30	--	??	3.02	3.41	+13%
Howard, IN	3.1	+0.74	+0.51	+0.74	-0.78	-0.40	-0.59#	--	+0.70	--	--	3.95	3.41	-14%
Iron, MI	3.1	-0.21	+0 14	--	+0.43	+0.08	+0.51	--	+0.10	--	--	3.71	3.41	-8%
Stephens, TX	3.1	-0.15	+0.47	--	-0.66	+0.32	-0.17#	--	+0.35	--	--	3.28	3.35	+2%
Throckmorton, TX	3.1	+0.21	-0.40	--	+0.17	+0.01	+0.18	--	--	--	--	3.28	3.34	+2%
Kit Carson, CO	3.1	+0.21	+0.70	+0.70	-0.02	-0.10	-0.06	--	--	--	--	3.74	3.30	-12%
Cottle, TX	3.1	-0.28	+0.33	--	+0.09	+0.01	+0.10	--	--	--	--	3.20	3.29	+3%

Table 1 (Cont'd.)

PREDICTIONS FOR 60 NON-METROPOLITAN COUNTIES

Line---→ County, State	S1	A4	A10	S2	B4	B8	S3	C6	C10	C14	C18	S5	Survey Result	Error
Dickey, ND	3.1	-0.49	+0.18	--	-0.08	-0.72	-0.40#	--	--	+0.40	--	3.10	3.24	+5%
El Dorado, CA	3.1	+0.78	+0.03	+0.78	+0.15	+0.92	--#	--	--	--	--	3.88	3.10	-20%
Charlotte, FL	3.1	+0.52	-0.06	--	-0.29	+0.45	+0.16	--	--	--	--	3.26	3.08	-6%
Palo Pinto, TX	3.1	+0.38	-0.82	--	+0.12	+0.08	+0.20	--	+0.25	--	--	3.55	3.08	-13%
Parker, TX	3.1	+0.29	-0.42	--	-0.02	-0.24	-0.13#	+0.30	--	--	--	3.27	3.05	-7%
Stevens, WA	3.1	-0.16	-0.34	-0.34	+0.32	-0.18	+0.14	--	+0.10	--	--	3.00	3.01	0%
Shoshone, ID	3.1	+0.31	-0.20	--	-0.45	-0.46	-0.45#	--	+0.25	--	--	2.90	2.93	+1%
Williams, ND	3.1	+0.03	+0.54	+0.54	-0.44	+0.05	-0.20#	--	--	--	--	3.44	2.87	-17%
Motley, TX	3.1	-0.42	+0.18	--	-0.33	+0.39	--#	--	--	--	--	3.10	2.86	-8%
Arenac, MI	3.1	-0.03	+0.11	--	-0.24	-0.42	-0.66	--	+0.30	--	--	2.74	2.83	+3%
Sutter, CA	3.1	+0.57	+0.34	+0.57	+0.33	-0.06	--#	--	+0.20	--	--	3.87	2.81	-27%
Webster, LA	3.1	-0.56	-0.19	-0.56	+0.13	+0.36	+0.49	--	--	--	--	3.03	2.76	-9%
Foard, TX	3.1	-0.22	-0.34	-0.34	+0.11	+0.23	+0.34	--	--	--	--	3.10	2.69	-13%
Union, OH	3.1	+0.44	+0.04	+0.44	-0.27	+0.05	-0.22	--	+0.40	--	--	3.72	2.67	-28%
Yuma, CO	3.1	-0.18	+0.44	--	-0.03	-0.04	-0.04#	--	--	--	--	3.06	2.63	-14%
Pasco, FL	3.1	-0.29	-1.24	-1.24	-0.51	+0.09	-0.42	--	--	+0.15	+0.10	1.69	2.60	+54%
Hernando, FL	3.1	-0.40	-0.78	-0.78	+0.07	+0.03	+0.10	--	--	--	--	2.42	2.56	+6%
Dickenson, MI	3.1	+0.03	+0.18	+0.18	+0.03	-0.04	-0.01	--	--	--	--	3.27	2.49	-24%
Bonner, ID	3.1	-0.06	-0.04	-0.06	+0.14	-0.10	+0.04	--	+0.20	--	--	3.28	2.46	-25%
Somervell, TX	3.1	-0.24	-0.94	-0.94	-0.24	+0.27	--#	--	--	--	--	2.16	2.33	+8%
Converse, WY	3.1	+0.23	+0.11	+0.23	-0.01	+0.02	+0.01	--	--	--	--	3.34	2.22	-34%
Hardin, TX	3.1	-0.24	-1.00	-1.00	-0.55	+0.01	-0.54	--	+0.15	--	--	1.71	2.18	+27%
Amador, CA	3.1	+0.53	-0.14	--	+0.55	+0.22	+0.77	--	--	--	--	3.87	2.14	-44%
De Soto, LA	3.1	-1.41	-0.84	-1.41	-0.38	+0.75	+0.37	--	+0.35	--	--	2.41	2.07	-14%
Comanche, TX	3.1	-0.23	+0.25	--	-0.50	-0.24	-0.74	--	--	--	--	2.36	2.04	-14%
Washington, CO	3.1	-0.10	-0.95	-0.95	+0.20	-0.58	-0.19#	--	--	--	--	1.96	1.91	-3%
Pinal, AZ	3.1	-0.54	-0.70	-0.70	+0.14	-0.12	+0.02	+0.10	--	--	--	2.52	1.82	-28%
Red River, LA	3.1	-1.87	-0.98	-1.87	+0.03	+0.29	--#	--	+0.40	--	--	1.63	1.80	+10%
Sabine, TX	3.1	-1.46	-1.02	-1.46	-0.44	-0.12	-0.56	--	--	--	--	1.08	1.36	+26%
Bienville, LA	3.1	-1.65	-1.57	-1.65	+0.05	+0.48	--#	--	--	--	--	1.45	1.28	-12%

The four "C" columns refer to the appropriate lines of Schedule C that give per capita spaces from additional resources. Since the research team was not in a position to obtain local information for use in executing Schedule C, an alternate procedure was used in which the survey printout was reviewed for the existence of additional resources as defined in the instructions of Section II and approximations made as would be done by someone at the scene. The research team tried to be as objective as possible in applying the criteria of Schedule C but the comparison is most suspect on this point. In two counties, King, Texas, and Ogemaw, Michigan, for example, there was evidence in the printout of transient or resort facilities that were not reflected in the census data on service employment. Question marks in Table 1 indicate uncertainty as to whether additional "summer resort" spaces should be added.

Column "S5" is the Final Estimate of per capita CCS from Line 5 of the Summary. The estimate is obtained by adding "S1", "S2", "S3", and those "C" columns that have entries. The next-to-last column gives the survey result in per capita CCS and the last column shows the "error." The measure of error is related to the operational situation in which a prediction is made prior to a survey. The question to be answered is how the survey result will compare with the prediction in terms of a percent deviation from the prediction. Thus, in Esmeralda County, Nevada, the survey result is one percent higher than the prediction, as indicated by "+ 1%". In Bienville Parish, Louisiana, on the other hand, the survey result is 12 percent lower than the prediction.

The errors in the final column can be regarded as a random variable. If the amount of error is totaled and divided by the number of counties, the average error is found to be zero. That is, the prediction technique is unbiased for this sample of sixty counties. In fact, there are 31 underestimates, 28 overestimates, and one zero error. A histogram of the error distribution suggests a normal population of errors in the 3000 counties from which the sample is drawn. Assuming this to be the case, the standard deviation of the sample can be computed and is found to be 18.25 percent. Given that the 60-county sample is a random (representative) sample of all non-metropolitan counties in the United States, the meaning of the standard deviation is as follows. If all such counties were surveyed, we would expect the survey results to be within plus or minus 18.25 percent of the prediction in about 68 percent of the cases. In the sample, there are actually 42 counties with errors less than 18.25 percent, exactly 70 percent of the cases. We would expect the survey results to be within plus or minus 36.5 percent of the prediction in about 95 percent of the cases. There are two counties, a little less than five percent of the sample, with errors exceeding 36.5 percent--Pasco County, Florida, and Amador County, California. The odds that a survey result would deviate from its prediction by more than 50 percent are about 1 in 100. There is one such case in this sample, Pasco County, where the survey result is 54 percent above the prediction.

The foregoing statements of accuracy and reliability are "point estimates" based on the reasonable assumption that the errors or unpredicted residuals are randomly distributed. Confidence inferences can also be made from the data. From the fact that sample means deviate from the population mean according to the variance divided by number in the sample, one can say that the average error, which is zero for this sample, has a standard deviation of 2.36 percent. Thus, it is highly unlikely that the prediction technique is biased more than plus or minus five percent when other county samples are considered. It can also be computed that the standard deviation, about 18 percent for this sample, will lie between 13 percent and 23 percent for the whole set of non-metropolitan counties at the 95 percent confidence level.

In developing this prediction technique, we used as a criterion of success that the survey result be within 25 percent, plus or minus, of the final estimate. The operational meaning of this accuracy criterion is that if the prediction were used in the assignment of people from risk areas the allocation per person could vary in actuality from 30 to 50 square feet, once survey results were in hand. In the Northeast Corridor, where an allocation of 20 square feet may be required, a prediction of such accuracy would assure that housing space would vary between 15 and 25 square feet when survey results were available. In California, where housing allocations could approach the fallout shelter standard of 10 square feet per person, use of a prediction of this accuracy would imply an actual space allocation of between 7.5 and 12.5 square feet. Shelter occupancy experiments have been conducted successfully at less than the lower figure.

If the standard deviation of the sample, 18.25 percent, is representative, one would expect survey results to be within plus or minus 25 percent of the prediction in about 83 percent of all cases. In the sample, 49 of 60, or 82 percent, are within 25 percent of the prediction. Moreover, the cases of larger error are balanced between underestimates and overestimates. Thus, cases where the method overestimates the survey result by more than 25 percent should occur less than nine percent of the time. In the sample, there are six cases of underprediction of per capita CCS by more than 25 percent and five cases of overprediction. However, all of the cases of overprediction are in the resource-poor counties on the second page of Table 1. While there remains the possibility that some of these counties were really not surveyed completely, this situation may be a weakness of the method that would limit its usefulness in areas of reduced space allotment, such as California.

Use of the "Census Estimate"

It has been noted that, with the exception of Schedule C, the method given in Section II can be easily performed on a computer with the use of a machine-readable copy of the 1972 County and City Data Book. The penalty for making an estimate of congregate-care space based only on

Schedules A and B, which can be called the "census estimate," is the failure to fully identify resource-rich counties. For example, if S1, S2, and S3 for Esmeralda County, Nevada, in Table 1, are added up, one obtains an estimate of 4.05 per capita CCS. The survey result is higher by 125 percent because Esmeralda County has a small population, 629 persons, and much mine space and a gambling casino that are not counted. This is, of course, an extreme example. Most resource-rich counties will not be so badly underestimated.

The error distribution of the census estimates has a mean error of about + 9%. That is, the predictions underestimate survey results by nine percent on the average. Moreover, the standard deviation of the distribution is nearly 28 percent. If the single county, Esmeralda, is omitted, the average underestimate is reduced to about seven percent and the standard deviation of the errors is reduced to a little over 23 percent. Thus, the census estimate as given in Section II is biased toward underestimation and is not as accurate as the final estimate using Schedule C.

The census estimate can be improved by some simple modifications to the numerical values given in Section II. The initial estimate should be increased to 3.30 and the fourth rule under Line 10 in Schedule B should be adjusted to enter only 25 percent of the Line 9 decrease when less than eight percent of the workforce works outside the county. If these changes are made, the census estimates are virtually unbiased, the mean error being within one percent of zero. For the 60-county sample, the standard deviation of the error distribution for the census estimate is about 25 percent. If Esmeralda County is omitted as an extreme case, the standard deviation of the errors becomes about 21 percent as compared with 18 percent for the final estimate, including Schedule C. This means that if the aforesaid adjustments are made, the computerized census estimates will have the property that survey results will be within plus or minus 21 percent of the estimate in about 68 percent of the cases, within plus or minus 42 percent in about 95 percent of the cases, and very rarely, as in the case of Esmeralda County, will the survey result deviate more than 63 percent from the estimate. Large errors will be predominately underestimates of resource-rich counties. In California and the Northeast Corridor where use of all housing resources is very important, this weakness in the census estimates can be very important.

1974 Complete Counties

White[3] identified 28 counties from the 1974 host area survey that appeared to be completely surveyed. Printouts of the survey data were not available for this study except for those counties located in the State of Colorado. Hence, comparison of the data with the prediction technique of Section II can be made only for the census estimate in most cases. For those counties for which a printout was available, a Schedule C estimate was made as well. The results are shown in Table 2.

Table 2

PREDICTIONS FOR 28 COUNTIES FROM THE 1974 HOST AREA SURVEY

County, State (Line----)	A4	A10	S2	B4	B8	S3	Census Estim.	Survey Result	Error	S4	S5	Survey Result	Error
Teller, CO	+0.00	-0.79	--	+0.31	-0.64	-0.33	2.97	8.27	+178%	+3.30	6.07	8.27	+36%
Gunnison, CO	+0.08	-0.08	--	+1.02	+0.44	+1.46	4.76	6.59	+38%	+0.90	5.46	6.59	+21%
Musselshell, MT	-0.62	+0.08	--	-0.23	+0.45	+0.22	3.52	5.53	+57%				
Alamosa, CO	-0.42	+0.56	--	+0.79	+0.05	+0.84	4.14	4.89	+18%	+0.15	4.09	4.89	+20%
Santa Cruz, AZ	-0.31	+1.41	--	+0.21	-0.02	+0.19	3.49	4.69	+34%				
Franklin, MA	+0.59	+0.10	+0.59	+0.08	-0.24	-0.16	3.73	4.48	+20%				
La Plata, CO	-0.05	+0.22	--	+0.32	+0.32	+0.64	3.94	4.28	+9%	--	3.74	4.28	+14%
Laurens, GA	-0.78	-0.28	-0.78	+0.08	+0.02	+0.10	2.62	4.18	+60%				
Hill, MT	+0.21	+0.52	+0.52	+0.25	-0.12	+0.13	3.95	3.96	0%				
Chaffee, CO	-0.18	+0.09	--	+0.19	+0.12	+0.31	3.61	3.90	+8%	+0.10	3.51	3.90	+11%
Fremont, CO	-0.44	-0.48	-0.48	+0.25	+0.16	+0.41	3.23	3.77	+17%	+0.25	3.28	3.77	+15%
Cochise, AZ	+0.08	-0.43	--	+0.74	+0.01	+0.75	4.05	3.72	-8%				
Pulaski, GA	-0.79	+0.50	--	+0.27	+0.62	--#	3.30	3.68	+12%				
Peach, GA	-0.63	-0.55	-0.63	+0.68	+0.51	--#	2.67	3.59	+34%				
Sussex, DE	+0.17	+0.07	+0.17	-0.10	-0.14	-0.24	3.23	3.46	+7%				
Bleckley, GA	+0.50	-1.21	--	+0.97	+0.09	--#	3.30	3.36	+2%				
Flathead, MT	+0.08	+0.28	+0.28	-0.05	+0.04	--#	3.58	3.26	-9%				
Graham, AZ	-1.12	-0.34	-1.12	+0.32	+0.15	+0.47	2.65	3.25	+23%				
Monroe, GA	-0.43	-0.75	-0.75	+0.06	+0.62	--#	2.55	3.16	+24%				
Glacier, MT	-0.72	+0.19	--	+0.52	+0.05	+0.57	3.67	3.09	-20%				
Toole, MT	+0.14	+0.27	+0.27	+0.31	+0.21	+0.52	4.09	3.00	-27%				
Choteau, MT	+0.69	-0.44	--	+0.10	-0.80	-0.70	2.60	2.67	+3%				
Saguache, CO	-1.60	-1.27	-1.60	-0.01	-0.74	-0.75	0.95	2.38	+151%	+0.60	1.35	2.38	+76%
Lewis, NY	-0.31	-0.45	-0.45	+0.21	-0.10	+0.11	2.96	2.36	-20%				
Greenlee, AZ	+0.41	-0.60	--	-0.60	-0.36	-0.24#	3.06	2.16	-29%				
Jones, GA	-0.41	+0.08	--	+0.01	+0.23	--#	3.30	1.55	-53%				
Williamson, GA	-1.32	-1.50	-1.50	+0.24	+0.13	--#	1.80	1.41	-22%				
Twiggs, GA	-2.08	-2.28	-2.28	+0.07	+0.45	--#	1.02	1.30	+27%				

Table 2 is similar in format to Table 1 with some modifications. Column S1, the initial estimate, has been omitted. For the census estimate, this value is 3.30; for the final estimates in Colorado, the initial estimate is 3.10, as in Section II. The census estimate (Schedules A and B only) has been introduced after column S3. It is the sum of 3.30 and the per capita CCS adjustments in columns S2 and S3. The survey results and the deviation from the census estimate are introduced at this point. For the Colorado counties, Line 4 of the Estimate Summary, S4, has been introduced. There follows the final estimate, S5, which is 3.10 plus S2, S3, and S4. The additional resources shown in S4 are described in a later paragraph. For convenience, the survey result for the Colorado counties is repeated, together with their deviation from the final estimate.

It can be seen from Table 2 that the final estimates for the seven Colorado counties consistently underestimate the survey results. Indeed, the average survey result is more than 25 percent over the estimate. The same bias can be seen in the census estimates where the average underestimate is about 21 percent. The "swing" of the calculation is about right. The estimates for Williamson and Twiggs Counties, Georgia, at the low end are in the right neighborhood, and, if the final estimates for Teller and Gunnison Counties at the top of the list are considered, these estimates are comparable to those near the top on the 1975 county list in Table 1. One partial explanation lies in the fact that the 1974 counties appear to be above-average with respect to the 1975 sample. The unweighted average per capita CCS for the counties in Table 2 is 3.64 whereas the average for Table 1 is 3.42. Similarly, the median value in the 1974 list is over 3.5 while the median in Table 1 is 3.3. A characteristic of the prediction technique--a deficiency as it were--is that it tends to underestimate the resource-rich counties in Table 1 somewhat and tends to overestimate the resource-poor counties on the second page of the table. The 1974 sample would merge preferentially into the upper part of the 1975 listing.

A statistical test indicates that the probability that the two samples were drawn from the same parent error distribution is only about 86 percent, rather lower than conventional tests of significance. The alternative is that either the 1974 survey was consistently more thorough than the 1975 survey or that some of the counties assumed to have been completely surveyed in 1975 may not have been. The conventional wisdom is that the 1974 survey was not as thorough as the later survey but there is no justification for this view in the data. Only eight of 28 counties in Table 2 are overestimated by the prediction technique derived from the 1975 survey results. One would need to adjust the initial estimate from 3.30 to 3.60 to eliminate the bias in the census estimates. If the counties in Table 2 were merged with those in Table 1, the initial estimate would have to be increased one- or two-tenths to eliminate the bias. Because we have not been able to work with final estimates for most of the counties in Table 2, this adjustment has not been made.

With respect to the Colorado counties where final estimates were made, some indication of the basis for the additional resources listed under column S4 is warranted. Teller County has usable mine space in the survey equivalent to nearly one space per capita. This would have been estimated on Line 6 of Schedule C. There is also one industrial facility that is large with respect to the population. Finally, the Cripple Creek and Victor areas of the county are renowned tourist attractions with many facilities of the resort type. This resource is not evident in column B8 and it is assumed that the early spring date of the census was in the off-season. A conservative estimate of the resort resources is about two spaces per capita. Gunnison County has some industry and a major ski resort at Crested Butte that is judged to be off-season.

The additional resources in Chaffee and Fremont Counties are industrial in nature. The additional resources in Alamosa and Saguache Counties are largely potato storage and processing facilities.

Additional 1975 Surveyed Counties

In addition to the 60 counties shown in Table 1, there are more than 100 other counties in which surveys were made in 1975. The data for some of these counties were obviously in error and uncorrectable. They were deleted from the data base. In some other cases, printout errors were identified and corrected or compensated for. In some instances, for example, overstatement of congregate-care space in a single or few buildings due to a key-punch error was treated as an additional resource in Schedule C for purposes of testing the prediction technique. A more detailed discussion of anomalies in the prediction data base will be found in Section V.

Ultimately, there were 19 counties identified in the residual group that could be classed as completely surveyed. These were counties that marginally failed the screening process described in Section V or that contained anomalies in their survey data which could be compensated for. Basically, these counties can be treated like the 60 counties in Table 1 except that we are less certain of the validity of the survey result. A comparison of the results of the prediction technique with the survey results for these 19 counties is shown in Table 3, which is similar in format to Table 1.

It can be seen that the predictions are in reasonable agreement with the survey results with a few exceptions. The survey results in nine of the counties are underpredicted; overpredictions occur in ten counties. The mean error is an overprediction of about four percent. The standard deviation of the errors is 25 percent, substantially higher than the 60-county value of 18 percent. These observations are consistent with the questions about the validity of some of the survey results for this group. If Floyd County, Texas, or Phillips County, Colorado, were

Table 3

PREDICTIONS FOR 19 ADDITIONAL COUNTIES FROM THE 1975 HOST AREA SURVEY

Line----- / County, State	S1	A4	A10	S2	B4	B8	S3	C6	C10	C14	C18	S5	Survey Result	Error
Stutsman, ND	3.1	-0.06	+0.53	--	+0.20	-0.28	-0.04#	+0.40	+0.10	+0.20	--	3.76	4.15	+10%
Niobrara, WY	3.1	+0.33	+0.44	+0.44	-0.07	+0.26	+0.19	+0.10	--	--	--	3.83	4.13	+8%
Carbon, UT	3.1	-0.06	-0.01	-0.06	+0.04	-0.02	+0.02	+1.85	--	--	--	4.90	3.87	-21%
Grant, NE	3.1	+1.44	-0.46	--	+0.08	+0.37	+0.45	--	--	--	--	3.55	3.69	+4%
Cole, MO	3.1	+0.58	+0.30	+0.58	+0.93	+0.03	+0.96	+0.15	+0.15	--	--	4.94	3.56	-28%
Hall, TX	3.1	+0.06	+0.03	+0.06	-0.15	-0.02	-0.17	+0.10	--	--	--	3.09	3.50	+13%
Schleicher, TX	3.1	+0.05	-0.56	--	+0.08	+0.11	+0.19	--	--	--	--	3.29	3.25	-1%
Weston, WY	3.1	+0.58	-0.08	--	+0.06	+0.22	+0.28	--	--	--	--	3.38	3.13	-7%
Benewah, ID	3.1	+0.27	-0.21	--	+0.28	-0.54	-0.13#	+0.10	--	--	--	3.07	3.03	-1%
Stonewall, TX	3.1	+0.50	-0.12	--	-0.52	-0.26	-0.78	--	--	--	--	2.32	3.02	+30%
Irion, TX	3.1	-0.50	-1.44	-1.44	-0.29	+0.05	-0.24	--	+0.50	--	--	1.92	2.94	+53%
Rolette, ND	3.1	+1.64	-0.53	--	+1.16	-0.48	+0.68	+0.10	--	--	--	3.88	2.90	-25%
Shackleford, TX	3.1	+0.42	+0.11	+0.42	-0.29	-0.30	-0.59	+0.30	+0.50	--	--	3.73	2.71	-27%
Divide, ND	3.1	-0.47	-0.41	-0.47	-0.18	-0.40	-0.29#	--	--	--	--	2.34	2.49	+6%
Kent, TX	3.1	-0.81	-0.21	-0.81	+0.04	--	+0.04	--	--	--	--	2.33	2.46	+6%
Floyd, TX	3.1	-0.39	+0.29	--	-0.24	+0.12	-0.06#	+0.50	--	--	--	3.54	2.35	-34%
Tyler, TX	3.1	-0.65	-0.32	-0.65	-0.01	+0.16	--#	--	+0.30	--	--	2.75	2.26	-18%
San Augustine, TX	3.1	-1.67	-0.70	-1.67	-0.34	+0.54	+0.20	--	--	--	--	1.63	1.83	+12%
Phillips, CO	3.1	+0.23	+0.32	+0.32	+0.05	-0.16	-0.06#	--	--	--	--	3.36	1.38	-59%

actually only partially surveyed, both the bias in the prediction and the increased variability could be accounted for.

1975 Partially-Surveyed Counties

Non-metropolitan counties that were surveyed only in part offer a number of difficulties in their analysis. If the total population of the county is used to determine the per capita CCS, the resulting value will be lower than that which would be obtained if the entire county were surveyed. Hence, at the least, a "survey population" must be established consisting of the residents in the parts surveyed. This is a difficult task and the results are inherently inaccurate. Details of the procedures used to establish a survey population for each partially-surveyed county are given in Section V.

There are several reasons why many counties were surveyed only in part. One group consists of counties that had some portion at blast risk according to DCPA risk criteria. Only the non-risk MCDs were to be surveyed in these cases. In some areas of the country, host counties were surveyed only to the extent that congregate-care space was needed for an assigned number of relocatees. When the predetermined amount of space was found, the survey was terminated to conserve resources. In some sparsely-settled counties west of the Mississippi, only the towns were surveyed to conserve resources. Barrens and rangeland having few resources were not visited. Each of these reasons for partial survey offered problems in determining a survey population. Nonetheless, a survey population was established for some 80 non-metropolitan counties in the 1975 survey. These estimates were made prior to the development of a prediction method. Thus, they form an independent basis for defining the survey result in terms of per capita CCS.

The survey results in 80 partially-surveyed non-metropolitan counties are compared with the final estimate according to the method of Section II in Table 4. In this table are given the 1970 census population and our survey population in addition to the final estimate, survey result, and their difference in terms of percent of the estimate. Before commenting on some obvious characteristics of this comparison, we should note that 60 percent of the errors are overestimates; that is, the survey results are less than the final estimates. Despite this, the mean error is in the opposite direction; the bias is about five percent below the survey result. Moreover, the dispersion of the error distribution has increased greatly. The standard deviation approaches 50 percent, double that in the previous comparisons. Prediction of congregate-care space in partially-surveyed counties is clearly less reliable than when the counties are completly surveyed. The primary reasons for this are discussed below.

Table 4

PREDICTIONS FOR PARTIALLY-SURVEYED COUNTIES

County, State	1970 Population	Survey Population	Final Estimate	Survey Result	Error
Harding, SD	1,855	403	0.94	6.55	+597%
Mackinac, MI	9,660	9,189	6.90	6.34	- 8%
Pend Oreille, WA	6,025	3,741	3.49	6.08	+ 74%
Baldwin, GA	34,240	32,484	4.32	5.78	+ 34%
Crook, WY*	4,535	2,355	1.91	5.44	+185%
Childress, TX	6,605	6,405	4.88	5.24	+ 7%
York, ME*	111,576	6,672	6.84	5.21	- 24%
Garza, TX	5,289	4,022	5.13	5.18	+ 1%
Reno, KS*	60,576	54,006	4.64	5.04	+ 9%
Pope, AR*	28,607	20,074	3.50	4.95	+ 41%
Aroostook, ME*	94,078	20,982	4.19	4.79	+ 14%
Dawes, NE	9,693	7,481	3.82	4.54	+ 19%
Lincoln, LA	33,800	30,284	4.79	4.49	- 6%
Essex, NY	34,631	29,062	4.63	4.26	- 8%
Cass, IN*	40,456	35,492	4.10	4.24	+ 3%
Williamson, TX*	37,305	36,000	2.80	4.22	+ 51%
Washington, ME*	29,859	4,597	2.60	4.18	+ 61%
Lowndes, MS*	49,700	40,576	3.94	4.10	+ 4%
Dickinson, KS	19,993	18,275	3.52	4.08	+ 16%
Box Butte, NE	10,094	6,862	3.16	4.01	+ 27%
Cloud, KS	13,466	12,112	3.50	3.97	+ 13%
Goshen, WY*	10,885	1,073	3.22	3.89	+ 21%
Lavaca, TX	17,903	3,299	2.27	3.82	+ 68%
Fall River, SD	7,505	6,034	3.81	3.77	- 1%
Hyde, SD	2,515	1,957	3.48	3.70	+ 6%
Sioux, NE	2,034	478	1.63	3.70	+127%
Independence, AR*	22,723	11,423	2.79	3.68	+ 32%
Chippewa, MI*	32,412	3,375	4.64	3.63	- 22%
Platte, WY*	6,486	1,992	2.97	3.60	+ 21%
Madison, OH*	28,318	19,146	3.65	3.60	- 1%
Howard, MO*	10,561	7,003	3.80	3.50	- 8%
Barber, KS	7,016	1,217	5.58	3.43	- 39%
Campbell, WY	12,957	10,265	2.90	3.42	+ 18%
Licking, OH*	107,799	93,551	4.12	3.42	- 17%
Searcy, AR	7,731	6,096	2.77	3.31	+ 19%
Barnes, ND	14,669	672	3.96	3.26	- 18%
Eddy, NM	41,119	35,112	3.86	3.23	- 16%
Erath, TX	18,141	17,141	3.62	3.21	- 11%
Marquette, MI*	64,686	24,548	3.87	3.16	- 18%

Table 4 (Cont'd.)

PREDICTIONS FOR PARTIALLY-SURVEYED COUNTIES

County, State	1970 Population	Survey Population	Final Estimate	Survey Result	Error
Fairfield, OH*	73,301	60,942	3.47	3.14	- 10%
Duchesne, UT	7,299	6,817	2.70	3.12	+ 16%
Cache, UT	42,331	22,333	3.44	3.11	- 10%
Garland, AR	54,131	47,731	3.49	3.03	- 13%
Burleigh, ND	40,714	35,496	3.99	3.00	- 25%
Seminole, OK	25,144	17,927	2.83	2.98	+ 5%
Stone, AR	6,838	1,866	0.91	2.96	+225%
Eddy, ND*	4,103	3,282	3.80	2.89	- 24%
Iosco, MI*	24,905	7,767	3.40	2.73	- 20%
Missoula, MT*	58,263	9,640	4.13	2.68	- 35%
Wayne, NC*	85,408	30,554	2.80	2.63	- 6%
Panola, TX	15,894	14,794	2.30	2.62	+ 14%
Strafford, NH*	70,431	37,735	3.57	2.51	- 30%
Custer, SD	4,698	4,562	4.23	2.42	- 43%
Sedgwick, CO*	3,405	2,810	4.03	2.42	- 40%
Mineral, NV*	7,051	1,056	5.01	2.42	- 52%
Woodruff, AR*	11,566	9,102	1.95	2.38	+ 22%
Clinton, NY*	72,934	31,702	3.72	2.38	- 36%
Clay, KS	9,890	8,773	3.81	2.32	- 39%
Valley, MT*	11,471	8,249	3.77	2.24	- 41%
Benson, ND	8,245	3,277	1.84	2.13	+ 16%
Sheridan, NE	7,285	5,603	3.65	2.12	- 42%
Perry, AR*	5,634	3,992	0.78	2.08	+167%
Burnet, TX	11,420	2,864	3.01	2.01	- 33%
Merced, CA*	104,629	81,875	3.50	2.00	- 43%
Arkansas, AR	23,347	18,818	3.12	1.99	- 36%
Wells, ND	7,847	3,937	2.74	1.87	- 32%
Chaves, NM*	43,335	2,237	3.51	1.87	- 47%
Hayes, TX	27,642	6,266	2.91	1.86	- 36%
Hot Springs, AR	21,963	18,769	2.08	1.76	- 15%
Claiborne, LA	17,024	14,159	2.28	1.52	- 33%
Shelby, TX	19,672	17,672	1.73	1.47	- 15%
Miami, IN*	39,246	13,626	2.80	1.47	- 48%
Ottawa, KS	6,183	5,866	2.83	1.45	- 49%
Vernon, MO*	19,065	3,060	3.40	1.38	- 59%
Yuma, AZ*	60,827	23,133	4.02	1.33	- 67%
Hunterdon, NJ*	69,718	48,992	4.08	1.33	- 67%
Georgetown, SC	33,500	30,669	2.24	1.23	- 45%
Twin Falls, ID	41,807	24,889	3.91	1.22	- 69%
Newton, TX	11,657	9,197	1.39	1.11	- 20%
Lewis & Clark, MT*	33,281	29,352	4.57	1.02	- 78%

*Partially at blast risk.

Extraordinary errors of over 100 percent in prediction occur in 5 of the 80 counties. All of these errors are underpredictions: Harding, South Dakota, 597%; Crook, Wyoming, 185%; Sioux, Nebraska, 127%; Stone, Arkansas, 225%; and Perry, Arkansas, 167%. These large errors are the reason that the mean error is a 5 percent underprediction although 60 percent of the errors are actually overpredictions of the survey result. All of these counties are small (less than 10,000 population), western counties with little urban population. Two of the counties, Crook and Perry, were partially surveyed because part of these counties were at blast risk. The others were surveyed in part for other reasons, most probably that the parts not surveyed were ranch country, barren of congregate-care resources. A review of these counties indicates that a complete survey would not have added significantly to the space found, with the possible exception of Perry County. Hence, it would have been more accurate to have considered these counties as completely surveyed even though no facilities were identified in some parts. If this were done, the revised errors are + 51% for Harding (still an underestimate), + 48% for Crook, - 47% for Sioux, - 11% for Stone, and + 89% for Perry. If these adjustments are made to the data, the mean error becomes - 8%, consistent with the dominance of overestimates, and the standard deviation is decreased to less than 40%.

Chenault[4] has observed that congregate-care space is preferentially located in the towns and that the amount of space seems to vary exponentially with city size. That is, the largest town in a county is likely to have a higher proportion of the space than the ratio of its population to the population of smaller towns in the county would suggest. While a detailed analysis of this phenomenon has not been found in the literature, the data in Table 4 would lend credence to the observation. It would appear that simply counting the population in the areas surveyed is not the best procedure. There are a number of additional cases in Table 4 where it appears that the survey population was underestimated for practical purposes. Likely instances include Pend Oreille, Washington; Box Butte, Nebraska; Lavaca, Texas; and Benson, North Dakota.

A related source of error in the estimates concerns those counties that were partly at blast risk. In many instances, the risk area contains the principal town or city. The prediction method would tend to overestimate the per capita CCS in the remaining portion in these cases. Counties partly at blast risk are indicated by an asterisk in Table 4. Those with the main town in the risk area are to be found in the lower part of the table and include Missoula, Montana; Mineral, Nevada; Clinton, New York; Merced, California; Chaves, New Mexico; Yuma, Arizona; and Lewis and Clark, Montana. The resources in the residual parts of these counties are greatly overestimated by a method based on county-wide census indicators.

Similarly, there are counties partly at blast risk that are located adjacent to major urbanized areas, particularly in the eastern part of

the country. It is the urban part of these counties nearby the metropolitan counties that is at blast risk. Again, a high proportion of congregate-care space is likely to be in the areas not surveyed. Examples in Table 4 include York, Maine, Licking, Ohio, Strafford, New Hampshire, and Hunterdon, New Jersey. The resources in these counties are overestimated by the prediction method.

In general, predictions of per capita CCS based on county-wide census indicators appear to be an unreliable means of estimating resources in parts of counties, unless most of the county is represented, including the major towns. Nearly 20 percent of the counties listed in Table 4 have estimate errors of less than 10 percent. About 40 percent of the survey results are within plus or minus 20 percent of the final estimates. In most of these cases, the survey population is a major part of the total population or the surveyed part of the county is representative of the whole. These conditions are not met in most of the partially-surveyed non-metropolitan counties in the 1975 survey. It might be possible, however, to develop ways to identify counties that have unbalanced characteristics, and to classify these so that modifications could be made to the prediction to account for them in large part.

Partially-Surveyed Metropolitan Counties

There are data from the 1975 host area survey on 27 metropolitan counties. Predictions for these counties are compared with survey results in Table 5. All but two of these counties were surveyed only in part. The two counties that were completely surveyed, Seminole, Florida, and Johnson, Texas, are indicated by asterisks. It can be seen that the estimate error in these two cases is quite small. This may indicate that the prediction method of Section II is quite adequate for whole metropolitan counties.

In general, the nonsurveyed parts of the other 25 counties were those minor civil divisions (MCDs) that were deemed at blast risk according to DCPA risk criteria. The population of the surveyed part can be readily established, at least as it was in 1970. Therefore, most of the error in the predictions comes about because the risk part of the county is the urbanized part adjacent to the central city and contains a preponderance of the congregate-care facilities. As a consequence, the prediction method overestimates the per capita CCS in the residual outlying part of these counties. Overestimates occur in 20 of the 27 counties or 80 percent of the partially-surveyed metropolitan counties. Most of the overestimates are sufficiently large that the average survey result is 25 percent below the prediction. Moreover, the dispersion in the data is as high as in the case of the partially-surveyed non-metropolitan counties discussed earlier.

Table 5

PREDICTIONS FOR METROPOLITAN COUNTIES

County, State	1970 Population	Survey Population	Final Estimate	Survey Result	Error
Delaware, OH	42,908	31,026	4.40	4.49	+ 2%
Miami, OH	84,342	70,888	3.32	4.46	+ 34%
Clark, NV	273,288	42,920	7.22	4.45	- 38%
Butler, OH	226,207	150,116	3.19	3.54	+ 11%
Seminole, FL*	83,692	83,692	3.07	3.23	+ 5%
Montgomery, OH	606,148	7,438	4.16	3.19	- 23%
Johnson, TX*	45,769	45,769	2.80	3.13	+ 12%
Pickaway, OH	40,071	27,786	3.14	2.99	- 5%
Riverside, CA	459,074	178,619	4.12	2.86	- 31%
Warren, OH	84,920	50,966	2.39	2.71	+ 13%
Placer, CA	77,306	59,411	3.91	2.68	- 31%
Berkeley, SC	56,199	23,846	1.39	2.66	+ 91%
Cleveland, OK	81,839	56,717	3.25	2.32	- 29%
Tarrant, TX	716,317	50,323	3.76	2.27	- 40%
Pinellas, FL	522,329	208,817	4.11	2.19	- 47%
Oklahoma, OK	526,805	28,549	4.42	2.06	- 54%
Clark, OH	156,946	49,235	3.68	2.05	- 44%
Greene, OH	125,057	8,202	3.10	1.98	- 36%
Saline, AR	36,107	34,262	2.03	1.93	- 5%
McClennan, TX	147,553	35,540	2.82	1.92	- 32%
Sacramento, CA	631,498	12,400	5.03	1.88	- 63%
Orange, FL	344,311	130,903	3.77	1.86	- 51%
Caddo, LA	230,184	48,954	4.42	1.84	- 58%
Spokane, WA	287,487	14,516	3.86	1.83	- 53%
Oneida, NY	273,037	6,630	4.03	1.66	- 59%
Dade, FL	1,267,792	113,975	4.25	1.55	- 64%
Bossier, LA	63,703	22,674	2.65	0.88	- 67%

*Completely surveyed.

The bias in the predictions could be reduced to small proportions by reducing the initial estimate from 3.1 per capita CCS to about 2.4 CCS but the standard deviation of the error distribution would be increased. Thus, a prediction based on county-wide census indicators is inherently unreliable unless most of the county is not at risk or the surveyed part is representative of the whole. This was also true in the 80-county sample of partially-surveyed non-metropolitan counties. Again, a modified prediction method might be a reasonable goal, particularly for counties where the survey boundaries are well-defined.

IV ANALYTICAL PROCESS

The purpose of this section is to describe the analytical process that led to the prediction method exhibited in Section II. It also serves to elaborate some characteristics and limitations of the prediction data base and of readily-available census and other information that might be used to predict the amount of congregate-care space that could be expected from a host-county survey. A more detailed discussion of the prediction data base used in the analysis is given in Section V.

Review of Prior Work

The first published attempt to predict congregate-care housing capacity that is relevant to this study was that of William L. White.[3] In this 1975 study, White evaluated the survey results from the 1974 host area survey, the first conducted by DCPA. He found that only counties that had been surveyed completely were useful for his analysis. The 28 counties shown in Table 2 (Section III) are those used in White's study. An important feature of the study was the attempt to predict congregate-care spaces through analysis of the "use class" data for the surveyed facilities. The use class code used in the survey is shown in Table 6. Each facility in the county printout was assigned one of these codes by the surveyor. A review of the printouts indicate many misassignments when the use class was compared with the facility name and a high use of the x9 codes labeled "Other." In order to use the data, White laboriously corrected the misassignments. He then attempted to compare the yield of congregate-care spaces with census and other indicators. Linear regression analyses were performed to establish the relationships. Figure 1 shows the result when retail sales data from the 1972 County and City Data Book is compared with the corrected data for code 53, Stores other than food stores, the category with the highest yield (12 percent of all spaces) for the 28-county sample.

The black dots in Figure 1 represent the individual 28 county results. The linear regression line is shown and its equation and statistics given. The line does not pass through the origin; thus, the equation indicates a small number (23.7) of spaces at zero retail sales. The slope of the line would not change significantly if it were forced to pass through the origin. The slope is 0.331 spaces per thousand dollars of retail sales or about 1 space for each $3000 of retail sales. The coefficient of determination, r^2, is nearly unity, indicating that retail sales is a very good predictor of space in this use class. On the other hand, there is a considerable scatter of the data points from the best-fit line.

Table 6

USE CLASS CODES IN HOST AREA SURVEY

Residential

11 Apartment/Hotel
12 Dormitory/Barracks
19 Other

Educational

21 Kindergarten/elementary school
22 Junior high/high/preparatory school
23 College/university
24 Business/professional/industrial school
25 Correctional schools
26 Library/museum
29 Other

Religious

31 Church/synagogue
32 Retreat/monastery/convent
39 Other

Government and Public Service

41 Hospital
42 Clinic
43 Utilities
44 Communication facilities
45 Offices
46 Jails/prisons/correctional institutions
47 Armories/monuments/memorials
49 Other

Commercial

51 Offices
52 Food stores
53 Stores other than food stores
54 Warehouses
55 Banks/financial institutions
59 Other

Industrial

61 Factory/plant/manufacturing center
62 Food processing plants
69 Other

Amusement/Meetings

71 Theatre/auditorium
72 Community Center
79 Other

Transportation

81 Railroad station/terminal
82 Bus station/terminal
83 Airport terminal
84 Airport hangars
85 Marine terminal
86 Automotive repair and storage
89 Other

Miscellaneous

99 Categories not covered above

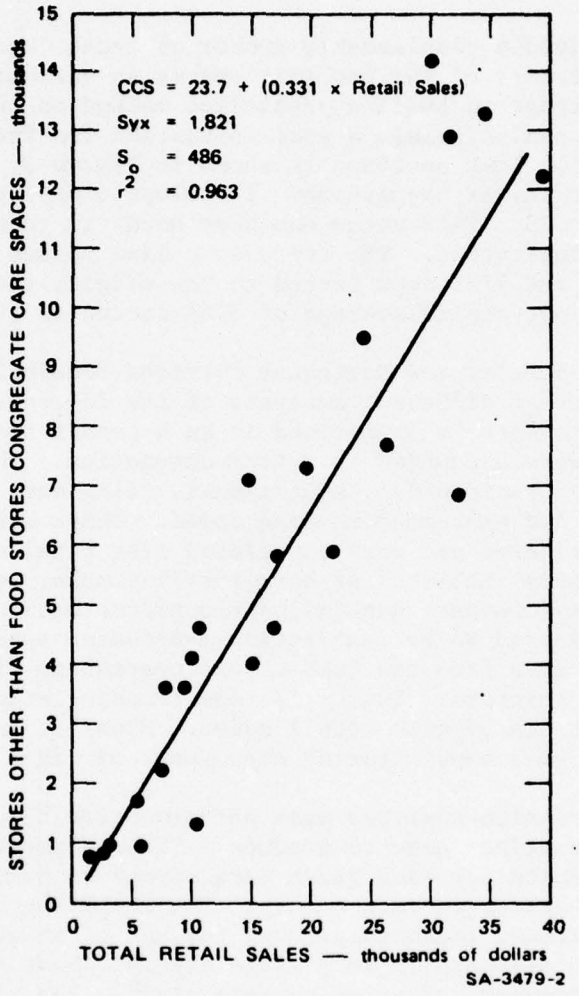


FIGURE 1 STORES OTHER THAN FOOD STORES CONGREGATE CARE SPACES AS A FUNCTION OF TOTAL RETAIL SALES

Figure 2, also from reference 3, shows the correlation of spaces in colleges and universities with enrollment data from the CCDB. The best-fit line is far from the origin and the coefficient of determination is very low, indicating that enrollment is a poor predictor for this use class.

White examined a considerable number of census and other data sources for useful indicators of the probable spaces in various use classes. Although his attempt to build a prediction method on this basis did not bear fruit, the analysis laid a good foundation for further work. A useful correlation that he found is shown in Figure 3, where total CCS is compared with county population. The slope of the best-fit line is 3.79 per capita CCS. This value has been used[1] in judging the feasibility of various hosting ratios. The regression line passes substantially below the origin. If the line were forced to the origin, the slope would approximate the per capita average of 3.64 quoted in Section III.

As a small part of the Northeast Corridor feasibility study, White conducted a somewhat different analysis of the 28-county sample from the 1974 survey. The work is summarized in an appendix to reference 1. The use class data were assembled into four categories. The first category consisted of most residential, educational, religious, government and public service, and amusement/meeting codes. These were considered to be population-oriented and were correlated with total county population. The second category consisted of dormitory/barracks, correctional schools, retreat/monastery/convent, and jails/prisons/correctional institutions. These were considered to be institutional-oriented spaces and were correlated with data from the CCDB on the percentage of the population living in group quarters. Commercial and transportation codes were combined and correlated with retail sales. Finally, industrial spaces were correlated with manufacturing employment as given in the CCDB.

Linear regression analyses were performed for the four groupings and the resulting equations used to produce a "four-element" prediction method in which the results for each group were summed to produce an estimate of congregate-care spaces in each county. The resulting estimates did not produce a significant improvement over the use of an average per capita CCS except for a decrease in very large errors. Some of the statistical findings were, however, of value to this study. The correlation of the first grouping with population was found to be very high. The commercial/transportation group had a high correlation with retail sales. However, spaces in group quarters and in industrial facilities were poorly correlated with the chosen census indicators.

Preparatory Analyses

One of White's recommendations was that the search for a usable prediction method be continued through use of the larger sample of county survey results produced by the 1975 host-area survey. The plan of work

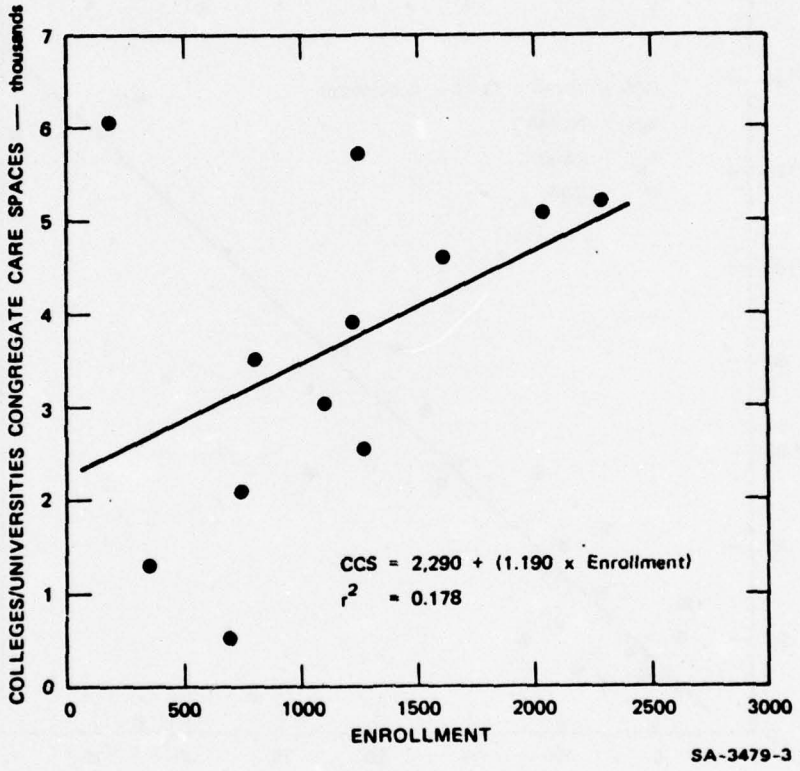


FIGURE 2 COLLEGES/UNIVERSITIES CONGREGATE CARE SPACES AS A FUNCTION OF ENROLLMENT

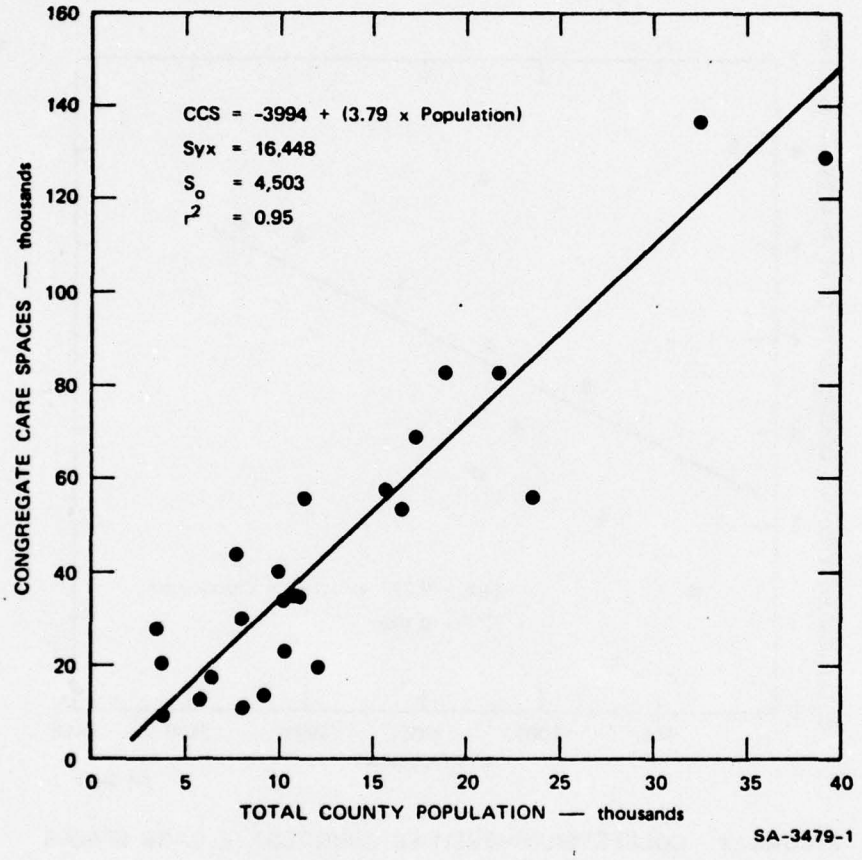


FIGURE 3 TOTAL COUNTY POPULATION AS A FUNCTION OF CONGREGATE CARE SPACES

for the analysis reported here called for the merging of the 1974 and 1975 survey results into a much larger data base that would hopefully lead to an improved prediction method. DCPA provided a computer printout of the 1975 survey results dated June 30, 1976. Not all of the survey data had been incorporated in this version but there was ample data to allow the preparatory analyses to begin. Later, the final printout was provided for those counties that were incomplete in the early version.

Initial review of the 1975 printout revealed most of the same problems encountered by White. Misassignment of use codes made the county-wide summaries questionable. The data base was too large, however, to make corrections feasible. Numerous errors, most attributable to key-punch mistakes, were found. These necessitated a careful review of the facilities listings. It is believed that the most important of these were corrected at various stages of the analyses. Difficulties in ascribing a survey population to partially-surveyed counties, as more fully described in Section V, and the questionable meaningfulness of the resulting per capita CCS, led to a decision to use only counties that appeared to be fully surveyed as the basis for development and test of candidate prediction techniques.

By the methods described in Section 5, a total of 40 counties were identified in the initial data base that were judged to be fully surveyed. These counties were assembled as the "base group" and used as the basis for development of predictive methods. The counties in the base group are listed in Table 7. The counties are ordered by survey result from highest per capita CCS to lowest. Esmeralda County, Nevada, has the highest per capita CCS (9.10) and Sabine County, Texas, has the lowest (1.36). The average county size is about 25,000 persons and the average yield in congregate-care spaces is about 92,000. The unweighted average per capita CCS is 3.64, very much like the average for the 28 counties in the 1974 survey. If a weighted average accounting for population size is computed, it is 3.62, as shown in parentheses in the table. This indicates negligible bias with respect to population size.

An initial analysis consisted of comparing the per capita CCS with the average value. Counties whose survey result was within plus or minus 25 percent of the average are indicated in the table by an asterisk. The operational significance of this criterion has been discussed in Section III. Twenty-five counties, or 62 percent of the sample, satisfy the criterion. Thus, simply predicting on the basis of the average would be adequate in nearly two-thirds of the cases. A more reliable method would need to account for the resource-rich and resource-poor counties at the extremes of the table.

A second step was to apply White's four-element method to the base group. Since the four-element method predicts total congregate-care spaces, the estimates were compared with the survey results shown in the second column of Table 7. The survey results were within plus or minus

Table 7

BASE GROUP COUNTIES

County, State	1970 Population	Congregate-Care Spaces	Per Capita CCS
Esmeralda, NV	629	5,727	9.10
Nye, NV	5,599	49,261	8.80
Collier, FL	38,040	219,948	5.78
Lincoln, NV	2,557	12,856	5.03
Luce, MI	6,789	32,450	4.78
San Luis Obispo, CA	105,690	501,416	4.74
Box Elder, UT	28,129	123,255	4.38*
Baylor, TX	5,221	22,283	4.27*
Manatee, FL	97,115	408,877	4.21*
Latah, ID	24,891	103,099	4.14*
Mason, TX	3,356	13,121	3.91*
Knox, TX	5,972	23,029	3.86*
Polk, FL	227,697	869,625	3.82*
King, TX	464	1,660	3.58*
Yuba, CA	44,736	159,938	3.58*
Dickens, TX	3,737	13,270	3.55*
Lincoln, WA	9,572	33,557	3.51*
Hood, TX	6,368	22,081	3.47*
Haskell, TX	8,512	29,175	3.43*
Hardeman, TX	6,795	23,252	3.42*
Ogemaw, MI	11,903	40,851	3.41*
Throckmorton, TX	2,205	7,373	3.34*
Kit Carson, CO	7,530	24,846	3.30*
Cottle, TX	3,204	10,540	3.29*
El Dorado, CA	43,833	135,696	3.10*
Charlotte, FL	27,559	84,805	3.08*
Palo Pinto, TX	28,962	89,093	3.08*
Stevens, WA	17,405	52,315	3.01*
Shoshone, ID	19,718	57,818	2.93*
Williams, ND	19,301	55,370	2.87*
Sutter, CA	41,935	117,665	2.81*
Foard, TX	2,211	5,957	2.69
Union, OH	23,786	63,486	2.67
Yuma, CO	8,544	22,436	2.63
Bonner, ID	15,560	38,244	2.46
Somervell, TX	2,793	6,520	2.33
Hardin, TX	29,996	65,440	2.18
Washington, CO	5,550	10,600	1.91
Pinal, AZ	67,916	123,806	1.82
Sabine, TX	7,187	9,750	1.36
Group Average	25,474	92,262	3.64 (3.62)

*Within Plus or Minus 25 Percent of Average

25 percent of the four-element prediction in 28 cases or 70 percent of the sample. Thus, there was a small improvement over the use of an average predictor. In general, however, the four-element method underpredicted the survey yield by an average of 18 percent. As a consequence, the method failed to predict the resource-rich counties, including some of the asterisked counties. In compensation, the method predicted satisfactorily many of the resource-poor counties at the bottom of the table. The predictions did not have sufficient "swing." That is, the counties with the highest per capita CCS were severely underpredicted and those with the lowest were seriously overpredicted. The range of predictions was considerably more restricted than the range of survey results.

As examples, the survey result for Esmeralda County, Nevada, was nearly four times the prediction, and the survey result for Sabine County, Texas was less than half the prediction. The error in the case of Esmeralda can be explained partially by the fact that the four-element method does not claim to account for what DCPA calls "special facilities," such as caves, mines, and tunnels. (See Instruction 4 in Section II for a more complete definition.) There is a large amount of mine space in the Esmeralda survey result. However, even if these spaces are deducted, the survey result is still underpredicted by a substantial margin.

The bias in the predictions by the four-element method for the base group could be eliminated by adding about two-thirds of a congregate-care space per capita to the computational formula. However, if this is done, adequate prediction of the resource-poor counties is lost without much improvement in the prediction of resource-rich counties. The method ends up predicting well the same counties that are predicted adequately by the average per capita CCS. The research team concluded at this point that the reason for this behavior lay in the use of the linear regression equations. That is, the four-element method predicts along the best-fit lines exemplified in Figures 1 through 3. Counties that are near the average in this sense are predicted well. Those that are far from the best-fit line are poorly predicted.

It seemed evident that the most likely avenue of improvement in predictive capability would be to assume initially that all counties were average in housing resources and to concentrate on the deviations from average characteristics that might predict both the resource-rich and the resource-poor counties. Additionally, it seemed desirable to eliminate population size from the calculation by phrasing it in terms of per capita CCS. Various manipulations of the base group data failed to show any significant trends with population size. Also, a prediction of per capita CCS is directly related to the concept of a hosting ratio.

The Basic Prediction Concept

The prediction concept underlying the procedure described in Section II was adopted at an early stage of the research. Time and effort

available did not permit the examination of other possible approaches. Thus, there is no guarantee that the approach adopted is the most effective or efficient one. It is also possible that an alternative concept could lead to a method with greater accuracy and reliability than our final procedure exhibits in the Section III discussion.

The basic concept was to assign every county an average per capita CCS, which would then be adjusted up or down based on census indicators. The coefficients used to convert census indicators to increments or decrements of per capita CCS would be chosen partly on the basis of data analysis and logical considerations and partly on the basis of empirical or pragmatic considerations--what worked best when compared with the survey results for the base group. The number of census indicators (and, hence, adjustments) to be included in the calculation of a predicted per capita CCS was also based partly on logical and partly on pragmatic considerations. A primary goal was to create a method that would swing away from the initial average to predict both the highest and lowest per capita CCS in the base group without disturbing the averageness of the majority of counties. In achieving this goal, it would be desirable to use the minimum number of adjustments possible.

Symmetric Trials

The initial set of trial solutions obtained using the basic concept was a set of symmetric adjustments to an initial "average." The first trial was a simple procedure involving three adjustments. A rounded version of the base group average, 3.5 spaces per capita, was used as the initial estimate. It was assumed that this initial estimate accounted for most of the spaces in the residential, educational, religious, public, commercial, and industrial sectors. Therefore, the adjustments were confined to the main resources believed to account for the wide variation between resource-rich and resource-poor counties. The first adjustment concerned resident institutions--colleges, reformatories, and military bases. The indicator used was column 16 in Table 2, CCDB, the percent of the population living in group quarters. This indicator was used despite the fact that White had found it unreliable in predicting specific use codes. For each county in the base group, the group average was subtracted from the percent tabulated in column 16. The result was multiplied by the empirical factor, 0.18, which implied that each percent of the population living in group quarters was equivalent to 0.18 CCS per capita. The adjustment would be positive or negative depending upon whether the county was above or below average in this regard. In a similar fashion, facilities serving transient or seasonal visitors were measured by two indicators--the percent employed in services and the amount of seasonal dwellings. The three adjustments were successively added or subtracted from the initial estimate to obtain a final estimate. When compared with the survey results, the initial trial seemed to indicate that the basic concept was promising.

Esmeralda and Lincoln Counties, Nevada, were still severely underestimated but Nye, Nevada; Collier, Florida; Luce, Michigan; and San Luis Obispo, California, were predicted adequately. The survey results for Nye and Collier were underestimated by 17 and 18 percent respectively while the survey results for Luce and San Luis Obispo were overestimated by 23 and 3 percent respectively. At the low end of the per capita CCS list, Hardin, Texas, and Washington, Colorado, were predicted within 25 percent and the survey result for Sabine, Texas, was only 37 percent below the estimate. These results indicated that the concept could generate swing away from the average that might be used to predict the counties at the extremes. The overall results, however, were not significantly better than using the average and suggested that a useful technique would be more complex.

Following this lead, a wide range of indicators from the 1972 County and City Data Book were investigated for their contribution to a predictive scheme. For those that seemed promising, national averages for non-metropolitan counties were computed to substitute for the base group average used initially. The way these national averages were obtained will be illustrated for the indicator in column 41 of the CCDB, the percent employed in services. From Table 1 of the CCDB, it is found that total census employment is 76,553,599 and the percent employed in services countrywide is 7.7 percent. This implies 5,894,627 employed in services in the United States at the time of census. From Table 3 of the CCDB, which concerns only metropolitan areas, it is found that employment in SMSAs is 54,034,345 and the percent employed in services is 8.0 percent. This implies 4,322,748 employed in services within metropolitan counties at the time of census. The differences, which apply to non-metropolitan counties, are 22,519,254 employment and 1,571,879 employed in services. The percentage employed in services is computed to be 6.98, which has been rounded to 7.0 percent for our purposes.

For a number of important indicators, Table 3 of the CCDB shows that the value for "all SMSAs" is not available. A search of the table revealed that the reason was the lack of information from one small metropolitan area. In these cases, the tabulated values were added to obtain an average for the metropolitan counties.

The second symmetric trial solution added retail sales in an attempt to measure variations in economic activity. Per capita money income was also added in a later version as an independent measure. Ultimately, the procedure used in the final version was hit upon. If both measures were above average, the largest was used. If both measures were below average, the most negative was used. This procedure minimized the weighting factor while conserving the necessary swing in the calculation. When the two measures were of opposite signs, it turned out best to ignore them and consider the county average in economic activity.

After successive trials, in the number of seven or eight, had given some indication of the sensitivity of the results to the key indicators and conversion factors, the analysis reached a dead end. The adequacy of predictions reached a plateau on which further innovations were of little use. The essentials of the situation were that swing could be generated more rapidly on the upward side of the average than on the downward and that there persisted a number of troubling anomalies in the performance of the method with respect to the survey results. A review of the analysis suggested several potential remedies. First, one could use a larger conversion factor for below-average performance than for above-average performance. This would increase the swing in the downward direction. There existed a straightforward rationale for the use of non-symmetric weighting or conversion factors. All of the census indicators being used in the prediction scheme were constrained in the below-average direction by the nearness of zero. For example, the lowest percent working in service activities was 2.8, which is 4.2 less than the average of 7.0. The highest in the sample was 19.4 percent, yielding an excess of 12.4 over the average. The greatest deficiency in per capita retail sales was $539 whereas the greatest surplus was $834. The greatest deficiency in per capita money income was $730 whereas the greatest surplus was $1685. All of this meant that there was a substantially greater opportunity to add to the initial estimate in above-average counties than there was to subtract in the below-average counties.

An alternative or additional change would be to lower the initial estimate, which had been based on the average per capita CCS. This change would reduce the bias toward the resource-rich counties at the price of increasing the underestimates made for some of these counties.

A third possibility was to recognize that some of the resources in the resource-rich counties, such as mines and caves, were not predictable by means of census indicators. By subtracting the per capita contribution of special facilities from the survey result, some of the results for resource-rich counties would be reduced dramatically. The prediction technique could be modified accordingly to the benefit of the resource-poor counties. If successful, this meant, of course, that the final estimate could not be made entirely on the basis of census indicators. The user of the technique would have to inquire at the local level to determine whether there were any special facilities that should be added.

All of the above remedies were undertaken. A revised set of survey results were developed by deleting space identified in special facilities. The initial estimate was reduced in increments of 0.1 per capita CCS. After some experimentation, all negative weighting factors were made twice as large as the positive factors. The trials made under these conditions were called "non-symmetric trials" because of the change in the weighting factors that converted census indicators to per capita CCS.

Non-Symmetric Trials

The deletion of special facilities from survey results and the use of non-symmetric conversion factors resulted in a sharp improvement in the accuracy and reliability of the predictions. There remained, however, a substantial number of poor estimates. Analysis in some detail of the facility printouts for these cases led to a final group of changes in the prediction technique.

One difficulty was that the census indicator, Percent Living in Group Quarters, was inadequate, as White had already pointed out, as a predictor of space in colleges, military installations, and other residence institutions. Most of these institutions were government-operated and supported. They were, for the most part, staffed by State employees. Trials using the percent employed in government in place of the group quarters indicator resolved most of these inaccuracies. A major private university would not be accounted for by the government employment indicator. Fortunately, most large private universities are not located in non-metropolitan counties. Nonetheless, private residence institutions could contribute in some cases. It was decided to treat these facilities as an additional resource similar to the special facilities.

A second difficulty was that the census indicators available were insensitive to the existence of unusual industrial facilities. A case in point was Box Elder County, Utah. The per capita CCS in this county was persistently underestimated during the trials. A review of the facilities listing for Box Elder County revealed that much of the underestimate could be accounted for by the substantial amount of space found in just one industrial concern, the Thiokol Corporation. None of the available indicators--manufacturing employees, value added, or establishments--would disclose this resource. Analysis of other counties indicated a similar pattern. Some small, rural counties were being underestimated because a single food processing plant provided significant CCS per capita. Of course, an average amount of industrial floor space was included in the initial estimate. But, unusual industrial resources would need to be treated much like special facilities and private residential institutions. A planner would need to establish their existence separately.

A third problem area concerned those counties with substantial resort facilities--hotels, motels, and camps, as well as the supporting services for a seasonal population increase. The census indicator, percent employed in services, is based on data obtained by the census in early April. This time is unfortunate in that it is usually the off-season for the tourist trade. Also, much of the seasonal employment needs are satisfied by non-residents of the county. Other possible indicators were found to be inadequate or unavailable for most non-metropolitan counties. Hence, it was ultimately decided to include seasonal resort facilities as additional resources to be investigated at the local level.

When these adjustments were made, only a few exceptional cases remained. Among these were a group of California counties in the foothills of the Sierra Nevada. These counties scored above-average not only on the basis of economic indicators but also on the basis of government and service employment. Yet no corresponding government or tourist facilities were evident in the survey results. Our data analysis indicated that it was highly unlikely that the survey would have omitted these facilities, if they had existed. A review of other census indicators in the CCDB disclosed that these counties, El Dorado and Sutter Counties, California, had an unusually high proportion of the work force working outside the county. The national average for non-metropolitan counties was 16 percent, whereas Sutter had 36 percent and El Dorado had 28 percent working outside the county. Apparently, many persons lived in the foothills and worked in Sacramento or other cities of the Central Valley.

A special study revealed that only when the percentage working outside the county was more than 50 percent greater than the average were major deficiencies in the imputed facilities noted. The study also revealed that some counties with an unusually low proportion working outside the county had more facilities than the government and services indicators predicted. Trials showed that accuracy was considerably improved if the percent working outside the county was taken into account. The rules described in Section II were ultimately developed as the best fit to the data.

Most non-symmetric trials were done in pairs, with the only difference being that retail sales per capita was used as the economic indicator in one and per capita money income used in the other. Comparison of these estimates soon revealed that neither indicator was completely adequate. In general, the highest indicator gave the best estimate if both were above average. Similarly, the lowest gave the best estimate if both were below average. In the cases where one indicator was positive and the other negative, the performance was mixed. It appeared best to ignore the difference in these cases and this rule was ultimately adopted.

At the completion of the non-symmetric trials using the base group data, all of the elements of the prediction technique of Section II had been defined. The initial estimate in use was 3.30 per capita CCS. Our criterion of success was that the survey result be within plus or minus 25 percent of the prediction. Table 8 shows the performance of the method with respect to the base group.

Table 8 shows the survey result, the prediction using only the census indicators (Schedules A and B), the final estimate, including Schedule C, and the error with respect to the prediction. Since the research team was not in a position to obtain local information for use in executing Schedule C, the data deleted in the process described above was added back into the final estimate as if local information had been available.

Table 8

BASE GROUP COMPARISON

County, State	Survey Result	Census Estimate	Final Estimate	Error
Esmeralda, NV	9.10	4.25	9.25	- 2%
Nye, NV	8.80	6.34	7.59	+ 16%
Collier, FL	5.78	5.13*	5.13	+ 13%
Lincoln, NV	5.03	4.34*	5.19	- 3%
Luce, MI	4.78	4.18*	4.78	0%
San Luis Obispo, CA	4.74	4.60*	4.60	+ 3%
Box Elder, UT	4.38	3.36	4.36	0%
Baylor, TX	4.27	4.35*	4.35	- 2%
Manatee, FL	4.21	3.80*	4.10	+ 3%
Latah, ID	4.14	4.60*	4.70	- 12%
Mason, TX	3.91	3.27*	3.52	+ 11%
Knox, TX	3.86	3.42*	3.52	+ 10%
Polk, FL	3.82	3.35*	3.85	- 1%
Yuba, CA	3.58	3.85*	3.85	- 7%
Dickens, TX	3.55	3.44*	3.54	0%
King, TX	3.55	3.09*	3.09	+ 15%
Lincoln, WA	3.51	4.20*	4.20	- 16%
Hood, TX	3.47	3.24*	3.24	+ 7%
Haskell, TX	3.43	2.63	2.83	+ 21%
Hardeman, TX	3.42	3.47*	3.92	- 13%
Ogemaw, MI	3.41	2.92*	3.22	+ 6%
Throckmorton, TX	3.34	3.48*	3.48	- 4%
Kit Carson, CO	3.30	3.97*	3.97	- 16%
Cottle, TX	3.29	3.40*	3.40	- 3%
El Dorado, CA	3.10	4.08*	4.08	- 24%
Charlotte, FL	3.08	3.46*	3.46	- 11%
Palo Pinto, TX	3.08	3.50*	3.75	- 18%
Stevens, WA	3.01	3.10*	3.20	- 6%
Shoshone, ID	2.93	3.07*	3.32	- 12%
Williams, ND	2.87	3.74*	3.74	- 23%
Sutter, CA	2.81	3.87	4.07	- 31%xx
Foard, TX	2.69	3.30*	3.30	- 18%
Union, OH	2.67	3.52*	3.92	- 32%xx
Yuma, CO	2.63	3.28*	3.28	- 20%
Bonner, ID	2.46	3.28*	3.48	- 29%xx
Somervell, TX	2.33	2.36*	2.36	- 1%
Hardin, TX	2.18	1.76*	1.91	+ 14%
Washington, CO	1.91	2.25*	2.25	- 15%
Pinal, AZ	1.82	2.62	2.72	- 33%xx
Sabine, TX	1.36	1.28*	1.28	+ 6%

*Within plus or minus 25 percent of Census Estimate.

xx Exceeds plus or minus 25 percent of Prediction.

Survey results that fall beyond plus or minus 25 percent of the final estimate are indicated in Table 8 by an xx after the error figure. There are four, with errors of 29%, 31%, 32%, and 33%. Since there are forty counties in the sample, the accuracy criterion is satisfied in 90 percent of the cases. Note, however, that all of the failures are overestimates of counties with below-average resources; that is, the survey results in these cases are more than 25 percent below the prediction. At this stage of the analysis, it was believed that some of these failures could be due to incomplete survey results and, in any event, the amount of error was not extreme.

Cases where the survey result is within plus or minus 25 percent of the census estimate are indicated by asterisks in that column. There are six failures; hence, the prediction based on census indicators alone "works" 85 percent of the time with this sample. Note that two of the final estimate failures in the lower part of the table are estimated with acceptable accuracy when Schedule C is not used. But there are now four underestimates of resource-rich counties in the upper part of the table that would be corrected by use of Schedule C.

1975 Test Group

While the analysis of the base group of counties was underway, a later printout of survey results (dated July 31, 1976) was received from DCPA. Use of the updated information permitted the identification of 20 more counties that appeared to have been completely surveyed. These 20 counties were considered a "test group." The prediction method developed using the base group was subsequently applied to the test group. Table 9 shows the results of applying the technique to the test group. The table format is identical with Table 8. Using the plus or minus 25 percent criterion for accuracy, there are five failures in the final estimate and six failures in the census estimate. That is, the final estimate is only 75 percent reliable and the census estimate only 70 percent reliable. The single "average" estimate of 3.30 is sufficient for 11 of 20, or 55 percent of the cases.

Generally, the prediction method developed on the basis of the "base group" appears to work less well when applied to the "test group." There are, however, some features of the test group that appear to account for most of the differences. The principal feature of the test group is that it consists primarily of resource-poor counties. The median value from the survey results is only about 2.8 whereas the median value for the base group is about 3.4 per capita CCS. Thus, if the two groups were merged, most of the test group would appear in the resource-poor part of the listing, which is where the failures occurred in the base group. This is also the region of the sample where the question of whether a complete survey was actually accomplished could be raised.

Table 9

TEST GROUP COMPARISON

County, State	Survey Result	Census Estimate	Final Estimate	Error
Albany, WY	5.57	4.69*	4.69	+ 19%
Whitman, WA	4.99	4.41*	4.41	+ 13%
Eastland, TX	4.71	3.26	3.61	+ 30%xx
Howard, IN	3.41	3.74*	4.44	- 23%
Iron, MI	3.41	3.81*	3.91	- 13%
Stephens, TX	3.35	3.21*	3.56	- 6%
Dickey, ND	3.24	3.10*	3.50	- 7%
Parker, TX	3.05	3.23*	3.53	- 14%
Motley, TX	2.86	3.30*	3.30	- 13%
Arenac, MI	2.83	2.64*	2.94	- 4%
Webster, LA	2.76	3.23*	3.23	- 15%
Pasco, FL	2.60	1.64	1.89	+ 38%xx
Hernando, FL	2.56	2.62*	2.62	- 2%
Dickenson, MI	2.49	3.47	3.47	- 28%xx
Converse, WY	2.22	3.54	3.54	- 37%xx
Amador, CA	2.14	4.07	4.07	- 47%xx
De Soto, LA	2.07	2.26*	2.61	- 21%
Comanche, TX	2.04	2.56*	2.56	- 20%
Red River, LA	1.80	1.43	1.83	- 2%
Bienville, LA	1.28	1.65*	1.65	- 22%

*Within plus or minus 25 percent of Census Estimate.

xxExceeds plus or minus 25 percent of Prediction.

The only exceptional failure in the test group was the underestimate by 30 percent of the survey result in resource-rich Eastland County, Texas. There were no such failures in the base group. Eastland County is an interesting case. It contains two large junior colleges, Cisco and Ranger, that contribute a greater-than-average amount of school and dormitory space. Although these institutions are State-owned, the percent of government employees in the county is below average. Even so, the county also possesses several large manufacturing plants and peanut processors having many storage facilities. As can be seen from comparing the census and final estimates for Eastland County in Table 9, the research team added 0.35 CCS per capita for this industrial activity. However, the survey printout would support adding as much as 0.55 units. Moreover, the county has an unusual amount of hotel, motel, camp, and supporting services around two large lakes. If the early spring census data on service employment does not reflect this activity or if it has expanded significantly since the 1970 census, a knowledgeable local official would probably have considered some of it an additional resource. Therefore, the underestimate of Eastland County is quite possibly due to an inadequate assessment of Schedule C by the research team.

Statistical Considerations

Following the application of the method to the test group and the observation that the method performed less well for this group, the two samples were merged into the 60-county sample discussed in Section III. As predicted, the test group expanded primarily the resource-poor group of counties. The error distribution for the 60-county sample was then subjected to a series of simple statistical tests. The method was found to be biased in the direction of overestimation of the survey results. This bias apparently came about from a failure to reduce the initial estimate sufficiently as the Schedule C category of additional resources was expanded. The census estimate was unbiased with an initial estimate of 3.30 but the final estimate, including Schedule C, required an initial estimate of 3.10. The inclusion of Schedule C, of course, always adds to the sample predictions.

A histogram of the error distribution showed a well-behaved bell shape that suggested that the parent distribution was normal. Assuming this to be the case, the variance of the sample was computed. As discussed in Section III, the standard deviation was found to be about 18 percent. These statistics are a more powerful measure of accuracy and reliability than our working criterion of plus or minus 25 percent, which was dropped in favor of the statistical statements. In particular, the standard deviation permits a stronger statement of the reliability of the procedure, including the low probability of very large errors in prediction. It also becomes a convenient criterion for comparing the method with other survey results, such as the 1974 survey and the results in partially-surveyed counties. It is of some interest that the error distribution did not show normal behavior until the technique was developed to near its final form.

Limitations of the Analysis

As has been noted, the level of effort available for this analysis did not permit exploration of all the possible approaches to the prediction problem. The basic concept adopted did lead to a technique with a reasonable performance, at least in non-metropolitan counties that have been completely surveyed. The performance of the method could probably be improved by "fine tuning" the weighting factors and decision rules. It is probably best to do this in conjunction with application of the method to the 1976 survey results when they become available.

A serious limitation of the prediction technique presented in Section II is that it does not depend entirely on readily-available census or similar data. The final estimate in resource-rich counties, which cannot be identified in advance, requires the identification and gross measurement of a number of additional resources at the local level. This process, described in Schedule C and its instructions, requires not only knowledge of the county but also has subjective aspects that may lead to error. The census estimate, which can be computerized, has an error distribution with a larger variance than does the final estimate. It may be possible to improve this estimate with further work.

The attempt in this study to associate the results of a partial survey with an appropriate survey population was unsuccessful. This failure is particularly important with respect to the host portions of metropolitan counties. Considerable additional work will be needed to develop a suitable prediction capability for partial surveys.

The use of indicators from the 1972 County and City Data Book in predicting the outcome of the 1975 survey undoubtedly accounts for some of the variance in the error distribution. The discrepancies are likely to increase when applied to later surveys until more up-to-date census information becomes available. The weighting factors in the method will undoubtedly require adjustment when new base information is available.

Errors and ambiguities in the survey data are discussed in the next section. They also contribute to the variance in the prediction errors. Better quality control in the survey process and increased reliance on error-checking routines in the computer-based records can probably reduce these difficulties in the future.

V PREDICTION DATA BASE

The analysis that led to the development of the prediction technique presented in Section II was based entirely on the results of the host area survey conducted in the summer of 1975. The data on this survey was provided in the form of a computer printout for each county surveyed. A complete set of printouts dated June 30, 1976 was provided in August of 1976 and formed the basis for most of the analysis. This printout was known to be incomplete; that is, not all of the data for some counties had been key-punched and taken up in the computerized file and the data for some surveyed counties was missing. Several months later, the final data for these counties was provided in the form of printouts dated July 31, 1976. Printouts were provided only for those counties where there was a change.

The purpose of this section is to record the data analysis that was performed to establish a prediction data base. It also serves to highlight some of the problems encountered in using the survey data for analytical purposes. Some of these difficulties contribute to the uncertainties in comparing the prediction technique with the survey results.

Initial Review

The first step in the data analysis was to review the county summaries. Each county printout consisted of a listing of all facilities surveyed in the county, followed by a "CRP HOST AREAS FACILITY SUMMARY AND ANALYSIS REPORT." It was this facility summary that was reviewed first. An example of the summary is shown in Figure 4. The example is for Shackelford County, Texas. Only part of the data shown in this summary was of interest in this study. On the topmost line are given the Standard Location Code or RSAC for the county (55TE), the county name, the 1970 census population of the county (3,323), the summary title, the date of the summary, and the page number in the overall file. Immediately below are the column headings. The two columns headed "CONGREGATE CARE" are pertinent to the analysis. The first column gives the number of facilities having congregate-care space and the second gives the number of spaces in these facilities. For Shackelford County, there are 61 facilities with 8,995 congregate-care spaces of 40 square feet each. If the county were completely surveyed, the spaces divided by the county population would indicate 2.71 spaces per capita. This number was added in pencil by the research team.

STATE = SHACKELFORD POP 3323 CRP HOST AREAS FACILITY SUMMARY AND ANALYSIS REPORT 6/30/76 PAGE 3625

		CRP FACILITY		NSS PF CAT 2+ CONGREGATE CARE				UPGRADABLE FALLOUT		UPGRADABLE BELOW GROUND		UPGRADABLE BSMT ONLY		UPGRADABLE ABVGRD ONLY	
		TOTAL	W/BSMT	FAC	SPACES	FAC	SPACES	FAC	SPACES	FAC	SPACES	FAC	SPACES	FAC	SPACES
TOTAL		61	4	0	0	61	8995	49	15756	4	1220	4	1220	45	14536
SOIL NGMT (CU YD)							(2.71)		15166		800		800		14366
HEAT-WATER-POWER															
HEAT		53	3	0	0	53	8328	44	14439	3	1004	3	1004	41	13435
WATER		60	4	0	0	60	8935	49	15756	4	1220	4	1220	45	14536
EMERGENCY POWER		1	0	0	0	1	216	1	865	0	0	0	0	1	865
MEDICAL		2	0	0	0	2	278	2	1115	0	0	0	0	2	1115
PHARMACY		1	0	0	0	1	90	1	360	0	0	0	0	1	360
BEDS, TOTAL--	61	5	0	0	0	5	433	5	1604	0	0	0	0	5	1604
COMMODES, TOTAL--	212	58	4	0	0	58	8859	48	15486	4	1220	4	1220	44	14266
OWNERSHIP		61	4	0	0	61	8995	49	15756	4	1220	4	1220	45	14536
FED GVT		0	0	0	0	0	0	0	0	0	0	0	0	0	0
STATE-LOCAL GVT		12	0	0	0	12	2393	9	4165	0	0	0	0	9	4165
PRIVATE-OTHER		49	4	0	0	49	6602	40	11591	4	1220	4	1220	36	10371
USE CLASS		61	4	0	0	61	8995	49	15756	4	1220	4	1220	45	14536
RESIDENTIAL		2	0	0	0	2	87	2	353	0	0	0	0	2	353
EDUCATIONAL		6	0	0	0	6	1617	3	2154	0	0	0	0	3	2154
RELIGIOUS		6	4	0	0	6	1037	6	2237	4	1220	4	1220	2	1017
GVT-PUBLIC SERV		8	0	0	0	8	1021	7	2240	0	0	0	0	7	2240
COMMERCIAL		31	0	0	0	31	4421	26	7546	0	0	0	0	26	7546
INDUSTRIAL		0	0	0	0	0	0	0	0	0	0	0	0	0	0
AMUSEMENT		0	0	0	0	0	0	0	0	0	0	0	0	0	0
TRANSPORTATION		8	0	0	0	8	812	5	1226	0	0	0	0	5	1226
OTHER		0	0	0	0	0	0	0	0	0	0	0	0	0	0
SPECIAL FACILITIES		3	0	0	0	3	1114	1	504	0	0	0	0	1	504
	SEATING CAPACITY	KITCHEN HOURLEMS													
DINING FAC	648	48	8	1	0	0	8	1119	7	2939	1	486	1	486	6	2453
RESTAURANT	208	20	3	0	0	0	3	97	3	391	0	0	0	0	3	391
CAFETERIA	250	18	2	0	0	0	2	217	1	504	0	0	0	0	1	504
SNACK BAR	190	10	3	1	0	0	3	805	3	2044	1	486	1	486	2	1558

FIGURE 4 EXAMPLE SUMMARY PAGE

The per capita CCS for Shackelford County is not unreasonable, being only about 20 percent below the average. However, about 15 counties were found with less than a half dozen facilities listed. These were set aside as probably in error.

Defining the Survey Population

The next step was to check the standard locations and place names in the facility listings for the remaining counties. Figure 5 is the first page of the facility listing for Shackelford County, Texas. The standard location (SL) is given in the first column. Most counties in the United States are divided into several "standard locations," which are areas based on the minor civil divisions (MCDs) existing at the time of the 1960 census. Surveyors were given a map showing the partitioning of the county into sequentially numbered SLs. It will be noted that all of the facilities on the first page of the Shackelford listing are in SL 1. For each county, the SL numbers listed were brought forward to the summary page. The appropriate volume of the Standard Location Code was then consulted to determine if all of the SLs in the county had been listed. Where this was not the case, the SLs listed were compared with the 1970 census maps to determine the 1970 census population in the part of the county surveyed.

In addition, place names were compared with census maps and the Rand-McNally mileage guide maps. It will be noted in Figure 5 that each facility is given two lines in the printout. The first line gives alphanumeric data to be discussed later. The second line shows the name of the facility, the address, the place name, the nearest cross street, and, if the facility had a basement, some shelter upgrading information. Shackelford County has two Standard Locations and both are exhibited in the complete listing. Hence, Shackelford was placed in the group tentatively identified as having been completely surveyed. However, all of the 61 facilities listed for Shackelford are located in the county seat, Albany, except for 12 facilities in the town of Moran. There are apparently no facilities in the smaller towns of Acampo, Ibex, or Sedwick or in the camping grounds noted on the highway maps. The lack of place names in some county listings raised the question as to whether the whole SL area had been surveyed.

The results of the SL analysis, including our provisional determination of a survey population, was submitted to the COTR with the request that the DCPA Regions be asked to comment on our conclusions. Meanwhile, those counties that appeared to have been completely surveyed were used in the initial part of the analytical process described in the previous section.

BEST AVAILABLE COPY

```
55TE - SHACKELFORD              PUP    3323  TYPE 1  CRP HOST AREAS FACILITY LISTING         6/30/76                     PAGE     3620

 STD   FAC   LAT-   LONG-    U  U S B USABL BSMT   S  USABLE ROOF  PS WALL      H P M P  B   C W O  B SC  P N NSS     U R O CONG   UPGRD UPGRD
 LOC   NO.   ITUDE  ITUDE    S  W F S BSMT  WALL   T  ABVGRD AREA  CP LENGTHS F O E H  E   O A I  U EA  F S PF 2+   P E I CAPF   FALL- SOIL
                             E  N   # AREA  EXP    Y  AREA         TA FR  SDE A W O A  D   M T N  R AP    S SPACE   S S S SPACE  OUT   RQMT
                                    T              S               N          T E I R  S   D E E  N T  C           R O T        SPACE CD YD
                                                                               R C M       S R              T       O N

0001 00001 324321 0991747 45 3 0 N     0  0000  3  10800  4500 00  60  75 Y N N N  0  4 2 N  0    0 0 N     0 3 0 1  270   360   346
     SHACKELFORD CO COURT HSE       MAIN ST               ALBANY          TX  NEAREST CROSS STREET - 2ND ST            BSMT--    0     0

0001 00002 324322 0991744 46 3 0 N     0  0000  2   2720  1600 00  40  40 Y N N N 10  4 2 N  0    0 0 N     0 3 0 1   68   136   165
     SHACKELFORD CO JAIL          E 2ND ST                ALBANY          TX  NEAREST CROSS STREET - WALNUT            BSMT--    0     0

0001 00003 324325 0991742 59 4 0 N     0  0000  1   4410  4900 00  70  70 N N N N  0  2 2 N  0    0 0 N     0 3 0 1  110   441   367
     VACUM ICE HOUSE          220 E 2ND ST                ALBANY          TX  NEAREST CROSS STREET - WALNUT            BSMT--    0     0

0001 00004 324323 0991742 31 5 0 Y  4869  6666  1   4869  6276 20  40 134 Y N N N  0  4 2 4  4   60 X N     0 2 0 1  243   486   232
     FIRST CHRISTINA CHURCH       S WALNUT                ALBANY          TX  NEAREST CROSS STREET - 2ND ST            BSMT--  486   232

0001 00005 324325 0991745 86 4 0 N     0  0000  1   4050  4500 99  45 100 N N N N  0  2 2 N  0    0 0 N     0 3 0 1  101     0     0
     LOWE ELECTRIC            312 E 2ND ST                ALBANY          TX  NEAREST CROSS STREET - WALNUT            BSMT--    0     0

0001 00006 324325 0991748 42 4 0 N     0  0000  2   4000  2500 00  50  50 Y N N N  0  2 2 N  0    0 0 N     0 3 0 2  100   200   225
     DENTIST OFFIC                E 2ND ST                ALBANY          TX  NEAREST CROSS STREET - MAIN              BSMT--    0     0

0001 00007 324326 0991748 49 4 0 N     0  0000  1   4950  5500 99  50 110 N N N N  0  2 2 N  0    0 0 N     0 3 0 2  123     0     0
     OLD ABA HOUSE THEATER        N MAIN ST               ALBANY          TX  NEAREST CROSS STREET - 2ND ST            BSMT--    0     0

0001 00008 324327 0991748 53 4 0 N     0  0000  1   9000 11250 00 125  90 Y N N N  0  6 2 N  0    0 X N     0 3 0 2  225   900   702
     BEN FRANKLIN             209 N MAIN                  ALBANY          TX  NEAREST CROSS STREET - 1ST ST            BSMT--    0     0

0001 00009 324328 0991748 52 4 0 N     0  0000  1  80400 10500 70  75 140 Y N N N  0  3 2 N  0    0 0 N     0 3 0 2 2010     0     0
     HOME TOWN GROCERY        201 N MAIN                  ALBANY          TX  NEAREST CROSS STREET - 1ST ST            BSMT--    0     0

0001 00010 324329 0991746 86 4 0 N     0  0000  1   3038  3375 99  80  50 Y N N N  0  2 2 N  0    0 0 N     0 3 0 2   75     0     0
     TACK AUTO SERVICE        320 E 1ST ST                ALBANY          TX  NEAREST CROSS STREET - 1ST ST            BSMT--    0     0

0001 00011 324329 0991747 59 4 0 N     0  0000  1    768   960 00  60  16 Y N N N  0  2 2 N  0    0 0 N     0 3 0 2   19    76   136
     CONOCO SERV ST                 1ST ST                ALBANY          TX  NEAREST CROSS STREET - MAIN ST           BSMT--    0     0

0001 00012 324330 0991748 59 4 0 N     0  0000  1   8640  9600 50 120  80 Y N N N  0  4 2 N  0    0 0 N     0 1 4 0  216     0     0
     ALBANY TIRE + SUPPLY         N MAIN                  ALBANY          TX  NEAREST CROSS STREET - 1ST ST            BSMT--    0     0

0001 00013 324346 0991726 59 4 0 N     0  0000  1   1523  1904 00  56  34 Y N N N  0  2 2 2 10 100 0 N     0 3 0 2   38   152   190
     LONGHORN RESTAURANT            HILL ST               ALBANY          TX  NEAREST CROSS STREET - ELLISON           BSMT--    0     0

0001 00014 324334 0991748 59 4 0 N     0  0000  1    792   990 00  45  22 Y N N N  0  2 2 2  6   48 0 N     0 3 0 2   19    79   125
     BELLS RESTAURANT               MAIN                  ALBANY          TX  NEAREST CROSS STREET - RAILROAD          BSMT--    0     0

0001 00015 324335 0991745 31 4 0 N     0  0000  2  13860  7700 40 110  70 Y N N N  0  4 2 4  4  120 0 N     0 3 0 2  346   693   525
     FIRST BAPTIST CHURCH           MAIN ST               ALBANY          TX  NEAREST CROSS STREET - 1ST ST            BSMT--    0     0
```

FIGURE 5 EXAMPLE OF FACILITY LISTING

In the process of searching the county listings for SLs and place names, over half of the listings set aside as probably in error were identified as actually part of the survey of a different county. Apparently, the county RSAC code had been misidentified, causing the error. In addition, some facilities were found in major listings that were miscoded from other counties. We believe that most such errors were found and adjustments made in the data for the affected counties.

Survey Policies

The analysis described above revealed many discrepancies and uncertainties regarding the extent of the host area survey. It was believed that some of these uncertainties could be resolved by a better understanding of the policies governing the survey operation. This aspect of the problem was investigated by the project consultant, Mr. Charles D. Kepple. His investigation included (1) office discussions with the DCPA Survey Project Officer and the Survey Technical Director in the Office of the Corps of Engineers, (2) a field visit to an ongoing survey operation in Pennsylvania in company with the COTR, and (3) a telephone interview with each of the Regional Survey Directors.

Kepple concluded that it was wrong to assume that a Standard Location had been completely surveyed if it appeared in the county listing because Regional policies in that regard were not consistent. The results of the telephone interviews with the Region staffs are reproduced in Appendix 1. Taken together with the analysis of the county listings, it was concluded that no county in DCPA Region 1 had been completely surveyed and that population estimates would be unreliable. Because Region 2 was not surveyed in the summer of 1975, this meant that no complete counties were available from the Northeast Corridor area of the country. The investigation also casts doubt on the reliability for analytical purposes of the survey results in several other DCPA Regions.

Test Procedures

The data analysis and investigations described above, coupled with progress in the prediction methodology, led the research team to conclude that it was important to include in the analysis only completely-surveyed counties. Therefore, an additional test routine was devised to provide an independent check on the prediction data base. Two parameters were determined from the totals on the summary pages for the candidate counties. The first was the number of facilities per thousand population. For Shackelford County (Figure 4), this number is 61 divided by 3.323 or 18.36. The second parameter was the number of spaces per facility. For Shackelford, this number is 8,995 divided by 61 or 147.46. These parameters were plotted as a function of county population and trends determined. The two parameters were also plotted against each other. Because of the large range in the county population size, these plots were cumbersome and are not reproduced in this report.

Despite the scatter of county data points, the trends in these parameters were clearcut. The number of facilities per thousand population decreases with increasing population size. For counties with less than 10,000 population, the parameter ranges between 25 and 50. Between 10,000 and 100,000 population, the points cluster between 15 and 35. The other parameter, spaces per facility, trends in the opposite direction; the parameter increases with increasing population. For populations over 10,000, the spaces per facility tend to be more than 100. In the smaller population size, the norm is less than 100. It is this reverse behavior that makes the per capita CCS more or less independent of population size. The smaller counties have large numbers of small facilities; the large counties have smaller numbers of larger facilities. In general, the facilities per thousand population are the better test of survey completeness.

In this regard, Shackelford County appears to be in a doubtful category. It has about 18 facilities listed per thousand population. Its cohorts in the 3000 to 4000 population class are Cottle County, Texas (3,204) with 31 facilities per thousand; Dickens County, Texas (3,737) with 45 facilities per thousand; and Mason, Texas (3,356) with 43 facilities per thousand. Because of its low facility count, Shackelford was not included in the base or test groups of complete counties. It will be found in Table 3 of Section III among the additional counties of which the research team is less certain.

A third parameter was also developed for screening purposes and applied to the data base, except for counties that were known to have been partially surveyed. The number of manufacturing, retail trade, and wholesale trade establishments given in the CCDB was compared with the number of commercial and industrial facilities listed on the county summary page. It can be seen from Figure 4 that the Shackelford listing has 31 commercial facilities and no industrial facilities. This is shown under the "USE CLASS" line items. Table 2 of the CCDB shows 3 manufacturing establishments (column 121), 61 retail trade establishments (column 132), and 10 wholesale trade establishments (column 159) in Shackelford County as of 1967. Thus, the survey printout accounts for less than half the census establishments. Many counties show more facilities than census establishments.

Survey Adjustments

It has been noted that some errors were detected in the assignment of facilities to counties and that these errors were corrected when found. Additional apparent errors were observed when the county facility listings were reviewed in detail in the course of preparing the Schedule C portion of the prediction method of Section II.

One type of apparent error seemed due to the miscoding of certain data entries. For example, the Shackelford County summary (Figure 4) shows a total of three special facilities with 1114 congregate-care spaces. As discussed in an earlier section, special facilities are mines, caves, tunnels, underpasses, and the like. Which facilities are classed as special facilities can be determined from the facilities listing. For example, in Figure 5 the first facility is the county courthouse. The data in the first line is interpreted as follows: The facility is located in SL 1. The facility number is 1. There follows the latitude and longitude of the facility. Next comes the use class. It is shown as 45, which relates to government offices, according to Table 6 of Section IV. The next column, "OWN," is an ownership code. Code 3 is local government ownership. The next column, "SF," is the special facilities code. "0" means the facility is not a special facility. There are no special facilities on the example page. When the listing is searched, the three special facilities are found to be public schools, coded as underpasses. The schools do not have basements and have no existing shelter space. While each school may have a pedestrian underpass associated with it. the capacities are appropriate to the aboveground parts of the schools.

The significance of this apparent error is that the research version of Schedule C simply adds the per capita amount of special facility space to the census estimate. Thus, in Table 3 of Section III, there is shown + 0.30 for Shackelford County under column C6. It is doubtful that these facilities are really additional to the schools already accounted for and therefore the final estimate is higher than it should be. No adjustments have been made for these probable errors.

A different kind of apparent error can be illustrated from Figure 5. Continuing the explanation of the data entries for the Shackelford County Courthouse, the next three columns after the "SF" column refer to basements. The "N" indicates the courthouse does not have a basement. Hence, there is no usable basement area and no basement wall exposure. The entry under "STYS" shows that the courthouse has three stories. Next, the usable aboveground area is given as 10,800 square feet. The roof area is shown as 4500 square feet. The next column codes the roof span, a technical consideration of no immediate interest. Then, the wall lengths for the courthouse are given as 60 feet on the front and 75 feet on the side.

Now, these key data entries can be checked for consistency. If one multiplies the front dimension, 60 feet, by the side dimension, 75 feet, one gets 4500 square feet, which agrees with the roof area. (If the building was L-shaped or had wings, the roof area could be smaller than the product of the dimensions, but it should not be larger.) If we multiply the roof area, 4500 square feet, by the number of stories, 3, we obtain 13,500 square feet as the total floor area in the courthouse.

The usable floor area given, 10,800 square feet, is exactly 80 percent of the total area, the standard factor for this use class. Thus, the data for the Shackelford County Courthouse is internally consistent. The third column from the left credits the courthouse with 270 congregate-care spaces, which is the usable aboveground area divided by 40 square feet per space (there is no basement).

If one runs down the congregate-care space column, one arrives at facility number 9, which is alleged to contain 2,010 spaces, making it by far the largest congregate-care facility in Shackelford County. This facility is the Home Town Grocery, a one-story building without a basement but with nearly eight times the usable floor space of the three-story county courthouse. One can check this data item by multiplying the front dimension, 75 feet, by the side dimension, 140 feet, to get the roof area, 10,500 square feet. But the usable aboveground area is nearly eight times the roof area. There appears to be an extra zero inserted at this point. Eighty percent of the roof area would be 8400 square feet. So, the proper usable floor area is either 8400 or possibly 8040 square feet, as a standard factor is not used in all cases. In any event, the actual capacity of the grocery is probably 210 spaces and not 2010 spaces.

When an apparent error of this sort was detected, some correction had to be made if the error was a large one. If detected early, the congregate-care total was adjusted. If detected late, as in the case of Shackelford County, the error was introduced into the Schedule C calculation as if it really were an additional resource. Thus, the + 0.50 in column C10 of Table 3 for Shackelford County is the consequence of the data entries for the Home Town Grocery. It can be seen that if this were not done, the final estimate for Shackelford County would have been quite close to the mark. Indeed, the census estimate is only eight percent above the survey result. Needless to say, we are most uncertain about the data for Shackelford County and others in Table 3.

Other cases were encountered in which the congregate-care space appeared to be understated. Multi-storied buildings were found with less usable floor space than the roof area. In Weld County, Colorado, there were 13 facilities in both school and commercial classifications that were listed as having no usable floor area and, hence, no congregate-care space. The roof areas would indicate a large amount of space in these facilities. Weld County and others with similar questionable entries were dropped from the prediction data base.

A listing of printout errors and suspected errors has been provided to the COTR separately. It is our opinion that error detection routines should be built into the computer program to flag many of these apparently faulty data items so that corrections can be made. Although the prediction data base used in this study was adjusted as well as possible, many errors were probably not identified.

Additional adjustments were made when the responses were received from the DCPA Regions. For example, Region 3 advised that Polk County, Florida, originally assessed as completely surveyed, was not completed in 1975. They identified another 1900 facilities and 200,000 spaces surveyed in 1976. These were added to the Polk County totals. The receipt of the final printouts for 1975 counties also permitted the adjustment of totals and the identification of additional counties that appeared to have been completely surveyed. Some of these counties were used in the test group and some are in the additional counties shown in Table 3. The survey populations for incomplete counties added by the final printout were not submitted to the Regions for comment.

VI CONCLUSIONS AND RECOMMENDATIONS

Discussion

The principal objective of this work was to develop a reliable procedure for estimating in advance of a survey the likely amount of congregate-care space to be found in a host area. To an extent, this objective has been accomplished, at least for whole non-metropolitan counties. Non-metropolitan counties are, of course, the predominant host area jurisdictions. The prediction technique presented in this report was developed through the use of logical and empirical analysis; that is, general economic and activity predictors were chosen to represent general kinds of housing resources and the conversion factors and decision rules were developed by fitting the prediction to the survey results from the prediction data base. It was a pleasant development to find that the accuracy and reliability of the technique could be summarized in terms of the statistics of the error distribution.

Although the basic prediction concept was based on the idea of regarding each county as average prior to any adjustments, the initial estimate had to be reduced below the average of the 1975 data base to produce an unbiased estimate. In part, this result was influenced by the fact that certain important resource elements had to be treated as additional resources because suitable census indicators could not be found. In this respect, the study failed to establish a procedure that did not depend on any local knowledge.

An unexpected outcome was that the prediction technique underpredicted the limited 1974 data base. The conventional view had been that the 1974 survey was less complete than the later surveys. If the 1974 and 1975 samples of "complete" counties had been merged, the initial estimate in the technique would have required an increase and there would have been a greater variation in the error distribution. The hypothesis that the 1974 and 1975 samples are comparable cannot be entirely rejected. The 28-county sample from the 1974 survey could have been biased toward resource-rich counties by chance, just as the 20-county test group from the 1975 sample appeared to be biased toward the resource-poor end of the spectrum. Moreover, the responses to the telephone interview recorded in Appendix 1 suggest that the pressure of inadequate funding may have led to survey policies that understate the housing resources in many of the 1975 counties that were assessed as having been completely surveyed. If this should be the case, the prediction technique is likely to underestimate the actual housing resources available in other counties.

When the prediction technique was compared with the survey results for the 60-county sample from the 1975 survey, the standard deviation of the error distribution was found to be about 18 percent. This implies that the chance is only about one in a hundred that the housing resources in a county will be found to be as much as 50 percent higher than the prediction or as little as half the prediction. It is our judgment that most of the error can be attributed to variations in the prediction data base. Some of the error can be attributed to the necessity to use census information that was gathered a number of years before the time of the surveys. And, of course, the constants in the prediction method may need some adjustments as well.

Any prediction technique of the kind explored in this study will be limited by the fact that the census indicators antedate the survey. Nothing can be done to improve this situation. A good deal could be done, however, to improve the prediction data base. One weakness was that DCPA Region 2 was not included in the 1975 survey. This could be corrected when the 1976 survey results are available. Another weakness was the prevalence of partial surveys in DCPA Region 1. Together, Regions 1 and 2 constitute the risk and host areas of the Northeast Corridor, where a reduced space allocation will be required in the available host areas even if every potential housing facility is identified in the host counties.[1] The development of the prediction method was not influenced by the inadequate data base in the Northeast. Since housing will be at a premium in this area, it would seem that the survey policy in Regions 1 and 2 should be identical with that in Region 7, where a similar situation exists.

To the extent that continued funding and other limitations dictate that the host area survey will take many years to accomplish completely, a reliable prediction technique will be invaluable in permitting regional, State, and risk-area planning to proceed in the interim. For this reason, it may be well worth the effort required to establish the validity of some of the survey results by allowing an independent contractor to conduct an on-site evaluation of the survey in key counties. There is precedent for this kind of research activity. The Research Triangle Institute did a number of sample surveys of shelter facilities during the 1960s.[5,6] The Institute for Defense Analyses and the Texas Department of Public Safety conducted an audit of survey results in some Texas counties in 1974.[3]

Many of the apparent errors in the prediction data base could be detected and corrected as a routine matter by greater emphasis on audit subroutines in the national computer program and corrective follow-up at the DCPA Regional Centers. These actions not only improve the data base for prediction purposes but also would be of great assistance to the host area planners.

Conclusions

1. An unbiased prediction technique has been developed that predicts the survey results in whole non-metropolitan counties with a standard deviation of about 18 percent.

2. It has not been possible to depend entirely on census indicators in making predictions of congregate-care space. The assessment of certain housing resources require obtaining information from the jurisdiction.

3. A major source of prediction error may lie in uncertainties regarding the survey data. Survey policies vary from Region to Region. Pressures to maximize the return from available survey effort may account for much of the variance in prediction errors.

Recommendations

1. The prediction technique presented in Section II of this report should be used until a more reliable or simpler method is developed.

2. The census estimate using Schedules A and B should be computerized using the modifications described in Section III for use in regional crisis relocation planning and in policy studies.

3. The prediction technique should be tested against the results of the 1976 survey and further improvements made.

4. Greater attention should be given to error audits of the host area survey information and corrective action undertaken to provide a better basis for planning.

5. DCPA Headquarters should review Regional survey policies for appropriateness and require explicit statement of areas or building types not surveyed for use in future planning.

6. DCPA should consider the advisability of conducting an independent evaluation of survey results in a sample of host counties as a basis for quality control.

7. The problem of estimating hosting resources in the non-risk parts of metropolitan counties should be made the subject of separate study.

APPENDIX 1

TELEPHONE SURVEY OF REGIONAL HOST AREA SURVEY POLICIES

TELEPHONE SURVEY OF REGIONAL HOST AREA SURVEY POLICIES

Question 1: Are all facilities surveyed in a no-risk SL once the SL has been entered?

Region I: (Ray Muise) Not necessarily. Depends on number of c.c. spaces needed by State planners in a given area. Also depends on survey costs and resources available. Sometimes one large community not at risk will provide needed spaces; it is more cost-effective in such a case to survey this population center only. A multiplier of needed spaces (safety factor) given by DCPA is used; it varies from year to year, but is usually 1.5 or 2.0.

Region II: (K. Edwards) Yes. Sometimes, however, if we have found enough c.c. spaces we establish priorities for surveying of facilities eliminating those which are not so habitable, such as: gasoline stations, foundries, warehouses, automobile repair shops, R.R. roundhouses, which are often not heated.

Region III: (N. Seidel) Coverage of facilities is limited to manpower available and based on number of spaces needed by State planners. A multiplier of two times needed spaces is used, varied according to physical obstacles found in area.

Region IV: (R. Meyers) Yes.

Region V: (Ron Morrison) Yes. Initially we had an agreement with the local DCPA officials to do 150 percent of required number of spaces as determined by planners. But once we started to survey we found it desirable to survey all facilities in an area.

Region VI: (R. Froseth) R. Kistner) Yes. NCP planners tell us the number of spaces required in a host area. When we reach that number we stop the survey in that area. If we do not reach that number we do all the facilities.

Region VII: (E. Kaufman C. Cook) Yes. We survey all buildings that are not private residences. Have not yet found an SL having excess c.c. spaces.

Region VIII: (H. Eck, R. Runnerstrom) Yes, except apartment houses, condominiums and private homes.

Question 2: How are the SL chosen for survey?

Region I: We are told by DCPA; also use Adagio printouts.

Region II: Select county(s) based on State planning requirements; then we do no-risk SLs. When an SL is split we arbitrarily classify it as "at risk," or "not at risk" then survey accordingly.

Region III: We chose SL for survey in population centers of greater density thereby reducing transportation costs. We also try to group the SL to reduce travel time between surveying operations. Most spaces are found in urban areas anyway.

Region IV: All SLs in a host county are surveyed.

Region V: The field survey personnel have a map showing risk and no-risk SLs will be surveyed and we follow their recommendations.

Region VI: State planners select them and set up priorities for surveying. Before this region had state planners RSEG selected SLs for host area survey.

Region VII: No exceptions. We do all SLs in the host area.

Region VIII: Determined by DCPA regional personnel and assigned to RSEG.

Question 3: Are all SLs, not at risk in a county, surveyed once the county has been entered:

Region I: We do not survey by SL, we survey by county.

Region II: In some counties we find no excess space, then we do the whole county. In others where we have excess space, we cut off heavy industry facilities. Sometimes we have a time restraint then we leave off and finish next year. Some counties are not designated as either "risk" or "host" in which case we do not survey that county. The Field Services Division tells RSEG which counties to survey.

Region III: No.

Region IV: Yes, all within the host area of the county.

Region V: Yes.

Region VI: In a rural area the survey unit is the county. In urban areas the SL is the survey unit. Once we find the needed spaces in either an SL or a county we stop surveying that unit.

Region VII: Yes, no exceptions.

Region VIII: Yes.

Question 4: Are all habitable facilities that meet the space criteria reported whether or not they are considered to be upgradable as fallout shelter?

Region I: Yes.

Region II: Yes.

Region III: Yes. Existing shelter space shown in NSS is not surveyed.

Region IV: Yes.

Region V: Yes.

Region VI: Yes.

Region VII: Yes.

Region VIII: Yes.

Question 5: Why are SL numbers and area names not always consistent with the NLC book?

Region I: Sometimes use the post office name which may be 50 miles away. Need to go to lat-long to identify. Lately we have required surveyors to use a sketch and reference the facility to the nearest crossroad.

Region II: Probably a mistake in reading the SL number. Other times the SL map is misread. The NLC refers to a township name--many times the citizens do not refer to the local by township, but use another designation. We use the common designation.

Region III: Question not asked of this respondent.

Region IV: Question not asked of this respondent.

Region V: Not asked.

Region VI: In rural areas we locate facilities by zip code which gives the post office name to the SL unit being surveyed.

Region VII: Not asked.

Region VIII: Not asked.

Question 6: What is the criterion for determining the total square feet of useable space in a basement?

Region I: Not asked.

Region II: Not asked.

Region III: Not asked.

Region IV: Not asked.

Region V: In the case of columns, partitions we reduce total space by 10 to 20 percent of inside dimensions. Sometimes surveyors use outside dimensions. The summer hire's judgment governs. Averages 80-85 percent of total area reported as useable space.

Region VI: Visual inspection floor space of all immovables subtracted from total area 20 percent of outside dimensional area is allowed for walls and partitions.

Region VII: All useable space reported. If we can move an object out, the space is reported. Space occupied by immobile objects is eliminated, like computers in a bank or communications equipment.

Region VIII: Individual surveyors judge what space of the total is occupied by immovable objects. This is subtracted from the whole and remainder reported.

Question 7: Are apartment houses surveyed? If so, is all habitable space reported for congregate care?

Region I: Not asked.

Region II: Not asked.

Region III: Not asked.

Region IV: Not asked.

Region V: This summer we have not surveyed any portion of an apartment house unless it was a real large one, or is in the N.S.S. as fallout shelter space. Apartment houses are usually in the residential areas anyway and we do not survey them unless a large public or commercial building exists.

Region VI: Yes. Only common area, such as: meeting rooms, laundry areas, storage space is reported.

Region VII: Yes. We list all space including private space in an apartment house. We do no condominiums. In Hawaii, however, we only reported the public space of apartment houses.

Region VIII: No.

Question 8: How do you ensure that every facility in a survey unit has been evaluated?

Region I: Not asked.

Region II: Not asked.

Region III: Not asked.

Region IV: Not asked.

Region V: The permanent personnel with the team estimate, in advance, the total number of facilities expected in a unit area--by talking to local C.D. and other officials. In this way they know what to expect. On the spot the extent of coverage is at the discretion of the summer hires.

Region VI: The map is divided into grids and squares are assigned to the summer hires. The student colors in a block when it is done. Supervisors check only N.S.S. listings.

Region VII: Coverage is logged on county maps. All roads and streets are covered completely. If a facility is missed it is not approachable by road.

Region VIII: Field supervisor knows about what to expect; he assigns areas to surveyors by quadrants and checks coverage against initial estimate.

REFERENCES

1. W. E. Strope, C. D. Henderson, and C. T. Rainey, The Feasibility of Crisis Relocation in the Northeast Corridor, Stanford Research Institute (December 1976).

2. U.S. Bureau of the Census, County and City Data Book, 1972 (A Statistical Abstract Supplement, U.S. Government Printing Office, Washington, DC (1973).

3. William L. White, Crisis Relocation Planning - Host Area Survey Analysis, Stanford Research Institute (July 1975).

4. Personal Communication with William W. Chenault, Human Sciences Research, Inc. (January 1975).

5. E. L. Hill et al., Analysis of Survey Data, Research Triangle Institute (February 1964).

6. E. L. Hill et al., Structural Characteristics of NFSS Buildings, Research Triangle Institute (June 1967).

DISTRIBUTION LIST

ORGANIZATION	NO. OF COPIES
Defense Civil Preparedness Agency Research Washington, D.C. 20301 Attn: Administrative Officer	50
Assistant Secretary of the Army (R&D) Washington, D.C. 20301 Attn: Assistant for Research	1
Chief of Naval Research Washington, D.C. 20360	1
Commander, Naval Supply Systems Command (0421G) Department of the Navy Washington, D.C. 20376	1
Commander Naval Facilities Engineering Command Research and Development (Code 0322C) Department of the Navy Washington, D.C. 20390	1
Defense Documentation Center Cameron Station Alexandria, Virginia 22314	12
Civil Defense Research Project Oak Ridge National Laboratory P.O. Box X Oak Ridge, Tennessee 37830 Attn: Librarian	1
Dr. John Billheimer Systan, Inc. P.O. Box U Los Altos, CA 94022	1
Dr. William W. Chenault Human Sciences Research, Inc. Westgate Research Park 7710 Old Springhouse Road McLean, VA 22101	1
Dr. Jiri Nehnevajsa Professor of Sociology Department of Sociology University of Pittsburgh Pittsburgh, PA 15213	1
Dr. Conrad Chester ERDA, Holifield National Laboratory P.O. Box X Oak Ridge, TN 37830	1
Mr. Walmer E. Strope Stanford Research Institute 1611 North Kent Street Rosslyn Plaza Arlington, VA 22209	1
Dr. Murray Rosenthal System Development Corporation 2500 Colorado Avenue Santa Monica, CA 90406	1
Mr. Don Johnston Research Triangle Institute P.O. Box 12194 Research Triangle Park, NC 27709	1
Mr. George Penebaker Department of Interior Defense Electric Power Administration - Room 6545 18th and C Streets, N.W. Washington, D.C. 20240	1

ORGANIZATION	NO. OF COPIES
Mr. Robert A. Harker Center for Planning and Research, Inc. 750 Welch Road Palo Alto, CA 94304	1
Mr. Richard K. Laurino Center for Planning and Research, Inc. 750 Welch Road Palo Alto, CA 94304	1
The Dikewood Corporation 1009 Bradbury Drive, S.E. University Research Park Albuquerque, NM 87106	1
Ohio State University Disaster Research Center 127-129 West 10th Avenue Columbus, OH 43201	1
Stanford Research Institute 333 Ravenswood Avenue Menlo Park, CA 94025 Attn: Mr. Robert M. Rodden	1
URS Research Company 155 Bovet Road San Mateo, CA 94402	1
Dr. Gerald Klonglan Department of Sociology and Anthropology Iowa State University Ames, IA 50010	1
General Leslie Bray Director Federal Preparedness Agency, GSA 18th and F Streets, N.W. Washington, D.C. 20405	1
Mr. Thomas P. Dunne Administrator, Federal Disaster Assistance Administration Room B-133, Department of HUD 451 7th Street, S.W. Washington, D.C. 20410	1
Mr. Howard McClennon President International Association of Fire Fighters 1750 New York Avenue, N.W. Third Floor Washington, D.C. 20006	1
General Manager International Association of Fire Chiefs 1329 18th Street, N.W. Washington, D.C. 20036	1
Mr. Bjorn Pedersen International Association of Chiefs of Police 11 Firstfield Road Gaithersburg, MD 20760	1
Mr. Ferris Lucas National Sheriff's Association Suite 320 1250 Connecticut Avenue, N.W. Washington, D.C. 20036	1
Mr. Gerald W. Collins, Executive Vice President National Defense Transportation Association 1612 K Street, N.W. - Suite 706 Washington, D.C. 20006	1
National Fire Protection Association Library 470 Atlantic Avenue Boston, MA 02210	1
National Bureau of Standards Disaster Research Coordinator Office of Federal Building Technology Center for Building Technology Washington, D.C. 20234 Attn: Mr. C. G. Culver	1
Command and Control Technical Center The Pentagon - BE 685 Washington, D.C. 20301	1

ORGANIZATION	NO. OF COPIES	ORGANIZATION	NO. OF COPIES
Mr. Louis V. Spencer Radiation Theory Section National Bureau of Standards Room C-313 -- Building 245 Washington, D.C. 20235	1		
National Academy of Sciences (JH-312) Commission on Sociotechnical Systems Committee on Fire Research 2101 Constitution Avenue, N.W. Washington, D.C. 20418	1		
Gov. Leo A. Hoegh Timpa Road Chipita Park, CO 80811	1		
The Council of State Governments Disaster Assistance Project 1225 Connecticut Avenue, N.W. Suite 300 Washington, D.C. 20036	1		
Attn: Mr. Hubert A. Gallagher	1		
Dr. Joseph E. Minor Texas Tech University Dept. of Civil Engineering P.O. Box 4089 Lubbock, TX 79409			

PREDICTION OF CONGREGATE-CARE SPACE IN NON-METROPOLITAN COUNTIES, (UNCLASSIFIED) Walmer E. Strope and Betty J. Neitzel:

Stanford Research Institute, 333 Ravenswood Avenue, Menlo Park, California 94025 February 1977 94 pp. Contract DCPA01-76-C-0298, Work Unit 2312G.

A procedure is presented for predicting the amount of congregate-care space existing in a non-metropolitan county prior to a facility survey. Estimates of accuracy and reliability are given, based on a 60-county sample from the 1975 DCPA host area survey. Suggestions are made for improvement. The procedure is not suitable for predicting the outcome of a partial survey.

PREDICTION OF CONGREGATE-CARE SPACE IN NON-METROPOLITAN COUNTIES, (UNCLASSIFIED) Walmer E. Strope and Betty J. Neitzel:

Stanford Research Institute, 333 Ravenswood Avenue, Menlo Park, California 94025 February 1977 94 pp. Contract DCPA01-76-C-0298, Work Unit 2312G.

A procedure is presented for predicting the amount of congregate-care space existing in a non-metropolitan county prior to a facility survey. Estimates of accuracy and reliability are given, based on a 60-county sample from the 1975 DCPA host area survey. Suggestions are made for improvement. The procedure is not suitable for predicting the outcome of a partial survey.

PREDICTION OF CONGREGATE-CARE SPACE IN NON-METROPOLITAN COUNTIES, (UNCLASSIFIED) Walmer E. Strope and Betty J. Neitzel:

Stanford Research Institute, 333 Ravenswood Avenue, Menlo Park, California 94025 February 1977 94 pp. Contract DCPA01-76-C-0298, Work Unit 2312G.

A procedure is presented for predicting the amount of congregate-care space existing in a non-metropolitan *county prior to a facility survey. Estimates of accuracy* and reliability are given, based on a 60-county sample from the 1975 DCPA host area survey. Suggestions are made for improvement. The procedure is not suitable for predicting the outcome of a partial survey.

PREDICTION OF CONGREGATE-CARE SPACE IN NON-METROPOLITAN COUNTIES, (UNCLASSIFIED) Walmer E. Strope and Betty J. Neitzel:

Stanford Research Institute, 333 Ravenswood Avenue, Menlo Park, California 94025 February 1977 94 pp. Contract DCPA01-76-C-0298, Work Unit 2312G.

A procedure is presented for predicting the amount of congregate-care space existing in a non-metropolitan *county prior to a facility survey. Estimates of accuracy* and reliability are given, based on a 60-county sample from the 1975 DCPA host area survey. Suggestions are made for improvement. The procedure is not suitable for predicting the outcome of a partial survey.